

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS MATEMÁTICAS
Departamento de Estadística e Investigación Operativa I



**MODELOS GENERALIZADOS DE RIESGOS
PROPORCIONALES PARA EL ANÁLISIS DE
SUPERVIVENCIA**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR**

Martha Lorena Avendaño Garrido

Bajo la dirección de la doctora

M^a Carmen Pardo Llorente

MADRID, 2013

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS MATEMÁTICAS

Departamento de Estadística e Investigación

Operativa I



Modelos Generalizados de Riesgos Proporcionales
para el Análisis de Supervivencia

Tesis Doctoral realizada por:

Dña. Martha Lorena Avendaño Garrido

Bajo la dirección de:

Dra. Dña. M^a Carmen Pardo Llorente

Madrid, Enero de 2013

MARÍA DEL CARMEN PARDO LLORENTE, PROFESORA DEL DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA I DE LA UNIVERSIDAD COMPLUTENSE DE MADRID

CERTIFICA:

Que la presente memoria titulada:

MODELOS GENERALIZADOS DE RIESGOS PROPORCIONALES PARA EL ANÁLISIS DE SUPERVIVENCIA

ha sido realizada bajo mi dirección por Martha Lorena Avendaño Garrido, Licenciada en Matemáticas y constituye su tesis para optar al grado de Doctor en Ciencias Matemáticas.

Para que conste, en cumplimiento de la legislación vigente y a los efectos oportunos, firmo la presente en Madrid a 29 de Enero de 2013.

María del Carmen Pardo Llorente

A mi madre,
Lorena

A mi esposo,
Carlos

Agradecimientos

Agradezco sincera y ampliamente a la profesora María del Carmen Pardo Llorente por haberme permitido trabajar con ella y poder realizar esta tesis bajo su dirección. Gracias por su orientación, esfuerzo y dedicación. Le doy las gracias, ya que como investigadora principal del proyecto MTM2009–06997 del Ministerio de Ciencia e Innovación, me ha permitido formar parte de él.

Expreso mi gratitud al Ministerio de Asuntos Exteriores y de Cooperación por haber financiado este trabajo doctoral mediante el programa MAEC–AECID.

Mi agradecimiento al Departamento de Estadística e Investigación Operativa I de la Universidad Complutense de Madrid por haberme concedido la oportunidad de realizar esta tesis en dicho departamento.

Gracias al profesor Laurent Bordes quien me orientó en el Department of Mathematics and its Applications en University of Pau.

Agradezco a Carlos porque su ayuda en \LaTeX hizo posible este documento. Pero sobre todo, por su constante apoyo en todos los aspectos de mi vida.

Finalmente, dedico esta tesis a mi madre, por su cariño incondicional y porque a pesar de la distancia siempre está presente en mi vida.

Índice general

Resumen	1
1. Preliminares	5
1.1. Conceptos Básicos en Análisis de Supervivencia	5
1.1.1. Tipos de Censura en los Datos	7
1.1.2. Funciones Importantes para Datos de Supervivencia	10
1.1.3. Función de Verosimilitud	13
1.2. Modelo de Cox de Riesgos Proporcionales	15
1.2.1. Función de Verosimilitud Parcial	19
1.2.2. Función de Verosimilitud Total	22
1.2.3. Simulación de Tiempos de Supervivencia	24
1.3. Modelos Semiparamétricos	26
1.4. Organización de los otros capítulos	34
2. Modelo Generalizado para Datos con Censura por la Derecha	39
2.1. Introducción	39
2.2. Generalización del Modelo de Riesgos Proporcionales	41
2.3. Interpretación del Modelo	43
2.4. Simulación de Tiempos de Supervivencia	45
2.5. Modelo Paramétrico	46
2.5.1. Definición del Modelo	46

2.5.2.	Estimador de los Parámetros y sus Propiedades Asintóticas .	47
2.5.3.	Aplicación a Datos Reales	50
2.6.	Modelo Semiparamétrico	55
2.6.1.	Estimación de los Parámetros	55
2.6.2.	Aplicación a Datos Reales	56
2.6.3.	Propiedades Asintóticas de los Estimadores de los Parámetros	61
2.6.4.	Estudio de Simulación	86
2.6.5.	Estimadores de la Función de Supervivencia	93
3.	Modelo de Cox para Datos Doblemente Censurados	103
3.1.	Introducción	103
3.2.	Modelo Latente para la Estructura de Datos	105
3.3.	Estimadores del Modelo de Cox	106
3.4.	Propiedades Asintóticas de los Estimadores de los Parámetros . . .	112
3.5.	Aplicación a Datos Reales	131
3.6.	Estudio de Simulación	135
4.	Inferencia para los Modelos	145
4.1.	Introducción	145
4.2.	Estadísticos Basados en Divergencias	148
4.3.	Distribución Asintótica de los Estadísticos	154
4.4.	Potencia Asintótica de los Estadísticos	160
4.5.	Comparación de los Estadísticos Propuestos	162
5.	Conclusiones y Futuras Líneas de Investigación	189
5.1.	Conclusiones	189
5.2.	Futuras Líneas de Investigación	191
	Summary	195

Apéndice I: Derivadas	199
A. Modelo Paramétrico Generalizado para Datos con Censura por la Derecha (Sección 2.5)	199
B. Modelo Semiparamétrico Generalizado para Datos Censurados por la Derecha (Sección 2.6)	201
C. Modelo de Cox para Datos Doblemente Censurados (Capítulo 3) . .	203
Apéndice II: Integración por Partes	205
Apéndice III: Bases de Datos	207
A. Cáncer de Laringe	207
Bibliografía	211

Resumen

El análisis de supervivencia es un conjunto de técnicas estadísticas para analizar datos que corresponden a tiempos de ocurrencia de algún evento. En biomedicina el evento de interés suele ser la muerte, esto es, el tiempo analizado es el tiempo de supervivencia.

Los principales objetivos del análisis de supervivencia son estimar la función de supervivencia, la cual es la probabilidad de observar el evento de interés después de cierto periodo y estimar la función de riesgo, la cual es la probabilidad de observar el evento de interés en un momento dado que se ha sobrevivido hasta ese momento. Los métodos no-paramétricos son útiles para obtener estas estimaciones, pero en muchos estudios médicos se tendrá información adicional del paciente que influirá en el tiempo de supervivencia, en este caso, el objetivo principal del análisis de supervivencia es el de proponer un modelo para la función de riesgo en presencia de variables explicativas.

El objetivo de esta tesis es doble. Por una parte, desarrollar nuevos modelos de análisis de supervivencia más flexibles que los existentes para datos con censura por la derecha, que es la censura que más se presenta en los datos reales y por ende, una de las más estudiadas en la literatura. Por otra parte, estudiar modelos de supervivencia conocidos para mecanismos de censura complejos. En ambos casos se desarrollará la teoría necesaria para llevar a cabo la inferencia estadística, y además, se realizarán estudios de simulación.

En cuanto al primer objetivo, se propone una nueva familia de modelos de riesgos proporcionales, cambiando la función nexa exponencial en el modelo de Cox por la función de distribución logística generalizada del Tipo-I. Se estudia este modelo generalizado de riesgos proporcionales para datos de supervivencia con censura por la derecha. Es necesario recalcar que la nueva familia de modelos tiene un parámetro más que el modelo tradicional de Cox, este parámetro extra aporta flexibilidad al modelo, de hecho, se obtienen conclusiones más precisas al interpretar el riesgo relativo asociado a este nuevo modelo que con el modelo de Cox. Se estudian las versiones paramétrica y semiparamétrica de esta nueva familia de modelos, se establece la consistencia y normalidad asintótica de los estimadores de los parámetros del modelo y se proponen diferentes estimadores de la función de supervivencia. En particular, para la versión semiparamétrica del modelo, se proponen tres estimadores de la función de supervivencia basados en el estimador de la función de riesgo base acumulativa. El estimador de la función de riesgo base acumulativa es el equivalente al estimador de Breslow en el modelo de Cox. Se lleva a cabo un estudio de simulación para muestras pequeñas y moderadas, tanto de los estimadores de los efectos de las covariables como de los estimadores de la función de supervivencia. Además, se ilustra este modelo con dos conjuntos de datos de supervivencia de cáncer, uno de cáncer de laringe y otro de cáncer de pulmón. También, se proponen estadísticos basados en la divergencia de Kullback-Leibler, de Jefreys y de Rényi para hacer contrastes de hipótesis sobre los efectos de las covariables y se obtiene tanto su distribución asintótica bajo la hipótesis nula que nos permite construir la región crítica del correspondiente contraste de hipótesis, como su potencia asintótica. Los estadísticos propuestos se comparan mediante un estudio de simulación con los estadísticos clásicos de Wald y cociente de verosimilitud.

Con respecto al segundo objetivo, se propone un modelo de regresión para da-

tos doblemente censurados. Los datos doblemente censurados han sido estudiados por diversos autores que han propuesto estimadores de la función de supervivencia para ellos, sin embargo no se han propuesto modelos con variables explicativas. En este trabajo, se modelan los datos doblemente censurados mediante variables latentes y considerando covariables. Usando las variables latentes se define un estimador de la función de riesgo base acumulativa. Una vez hecho esto, se estiman los parámetros del modelo de Cox maximizando la verosimilitud “profile”. Además, se establece la consistencia y la normalidad asintótica de estos estimadores. Posteriormente, mediante simulación se estudia el comportamiento de los estimadores para muestras pequeñas o moderadas. También, se ajusta el modelo de Cox a datos reales doblemente censurados, dichos datos son de un estudio sobre la presencia de SIDA en mujeres infectadas con el VIH. Finalmente, se proponen estadísticos basados en la divergencia de Kullback–Leibler, de Jeffreys y de Rényi para hacer contrastes de hipótesis sobre el efecto de las covariables. Se establece la distribución asintótica de los estadísticos propuestos, para así construir la región crítica del contraste de hipótesis correspondiente. También se obtiene la potencia asintótica de los estadísticos. Por último, se realiza un estudio de simulación para comparar entre sí los estadísticos propuestos y los estadísticos clásicos de Wald y de cociente de verosimilitud, ya que a pesar de que asintóticamente estos estadísticos tienen la misma distribución, para muestras pequeñas o moderadas su comportamiento no tiene porque ser el mismo. Los estadísticos se comparan en cuanto a potencia y tamaño.

Capítulo 1

Preliminares

1.1. Conceptos Básicos en Análisis de Supervivencia

El análisis de supervivencia es un conjunto de métodos, técnicas y/o procedimientos estadísticos para analizar datos que corresponden a tiempo de ocurrencia de algún evento en particular, llamado tiempo de supervivencia. Algunas veces se puede determinar un tiempo origen y los datos corresponden a la duración de algún fenómeno. Estos datos pueden provenir de diversos campos como la medicina, las ciencias biomédicas y la epidemiología entre muchos otros. Hoy en día, una de sus principales ramas de aplicación es la medicina, donde el tiempo origen puede ser el momento de diagnóstico de alguna enfermedad o el momento de inicio de algún tratamiento para ésta y el evento de interés la muerte, es decir, literalmente el tiempo de supervivencia. Sin embargo, no todas las aplicaciones son de tipo fatal, por ejemplo, el evento de interés puede ser la remisión de la enfermedad y, en este caso, el tiempo de supervivencia es el momento en el cual el individuo es declarado sano.

El objetivo del análisis de supervivencia es poder predecir la probabilidad de

que ocurra el evento de interés para algún individuo, esto es, la probabilidad de sobrevivir, o de cura, dependiendo el tipo de aplicación, tomando como base la información obtenida de pacientes anteriores quienes proporcionan los datos de supervivencia.

Cabe mencionar que el análisis de supervivencia no sólo se aplica en ciencias biológicas, sino también en sociología, marketing, criminología, ingeniería, etc. Por ejemplo, en sociología se puede estudiar la duración del primer matrimonio o en ingeniería el tiempo de vida de algún componente electrónico. Sin embargo, en esta tesis todos los ejemplos hacen referencia a aplicaciones biomédicas.

Es necesario decir que los datos de supervivencia no son simétricos. En general, un histograma de tiempos de supervivencia de un grupo de individuos homogéneo tenderá a ser positivamente sesgado, es decir, tendrá una “cola” larga a la derecha del intervalo que contiene el mayor número de observaciones.

Una segunda característica que hace que los métodos comunes de análisis estadístico de datos sean inadecuados para este tipo de datos es que los tiempos de supervivencia son frecuentemente censurados. Se dice que el tiempo de supervivencia de un individuo es censurado cuando el evento de interés no ha sido posible observarlo en ese individuo. Por ejemplo, supóngase que después de incluir a un paciente en un ensayo clínico éste cambia de residencia, la única información disponible de ese paciente será la de la última cita o el último chequeo. Esto hace que se deban aplicar técnicas estadísticas particulares para este tipo de datos. En general, si un dato presenta censura será llamado *dato censurado* y si no la presenta se dirá que es un *dato completo*. El término censura fue utilizado por primera vez por Hald (1949). Existen diversos tipos de censura en los datos, los cuales se describen a continuación.

1.1.1. Tipos de Censura en los Datos

Cuando se formula un modelo para datos de supervivencia se deben establecer las características de los datos a los cuales se ajustará dicho modelo, por ello es necesario conocer de antemano el tipo de censura que tendrán los datos. Los datos censurados se pueden clasificar en datos con censura por la derecha, por la izquierda o de intervalo y estas censuras a su vez se pueden subclasificar en varios tipos o casos. Cabe mencionar que, en la literatura existen diversas clasificaciones y generalizaciones de la censura. A continuación se describen los tipos clásicos de censura que se asumen en los datos a los cuales se aplican los modelos de análisis de supervivencia.

De ahora en adelante, se denotará por T la variable aleatoria que indica el tiempo de supervivencia.

Censura por la derecha

Una observación es censurada por la derecha en un tiempo C si el tiempo de supervivencia T no se conoce, pero se sabe que es mayor que C . Sea T_0 el tiempo origen y T el tiempo de supervivencia observado, tal que $T = T_0 + T'$ o en cambio, si se hace una observación C' unidades de tiempo después del tiempo T_0 , en el tiempo $C = T_0 + C'$, y se sabe que hasta el momento C no se ha presentado el tiempo de supervivencia pero no se puede volver a hacer otra observación. En este caso, un dato puede expresarse por el vector aleatorio (X, δ) donde $X = \min(T, C)$ y $\delta = \mathbb{1}_{\{X=C\}}$ es la variable indicadora de censura.

Por ejemplo, en aplicaciones biomédicas, en particular en ensayos clínicos donde el tiempo de interés es el momento de la muerte. En este caso, un grupo de individuos entran en el ensayo clínico en determinado momento, quizá diferente momento para cada uno y se define el tiempo que durará el estudio. Al final del estudio algunos individuos continuarán con vida, esto es, el evento de interés no será observado. También puede ocurrir que algunos individuos abandonen el estudio

y por tanto, sólo se tiene información de ellos hasta determinado momento. Otra razón por la cual se puede tener censura por la derecha es que el individuo muera por causas ajenas a la enfermedad de estudio.

Censura por la izquierda

De forma similar a un dato con censura por la derecha, se puede expresar un dato con censura por la izquierda con el vector aleatorio (X, δ) donde $X = \max(T, C)$ y $\delta = \mathbb{1}_{\{X=C\}}$ es la variable indicadora de censura.

La censura por la izquierda es menos común que la censura por la derecha. Se puede encontrar en un estudio sobre el primer consumo de alguna droga. Algunos individuos recordarán la edad a la cual consumieron por primera vez y algunos otros no lo recordarán o no querrán dar esa información, sin embargo se sabe que la han consumido, por tanto se tiene una censura por la izquierda en el momento actual.

No obstante, en un estudio clínico se podrán encontrar datos con censura por la izquierda y datos con censura por la derecha, en este caso los datos de supervivencia serán **doblemente censurados**. Diversos autores como Turnbull (1974), Patilea y Rolin (2006) y Julià y Gómez (2011) han propuesto estimadores de funciones de supervivencia para datos de supervivencia doblemente censurados. En el Capítulo 3 de este trabajo, se definirán detalladamente los datos doblemente censurados, ya que se considera esta estructura de datos mediante el uso de variables latentes como en Patilea y Rolin (2006).

Censura de intervalo

Se presenta cuando el tiempo de supervivencia T no es observado, pero se puede ubicar en algún intervalo de tiempo a partir de un tiempo de observación O o tiempos de observación O 's. Existen dos tipos de esta censura, estudiados por Groeneboom y Wellner (1992) y Sun (2005).

- Caso I. *Datos de estado actual*. Sea O un tiempo de observación y T el tiempo

de supervivencia. Una observación consiste en el vector aleatorio (O, δ) con $\delta = \mathbb{1}_{\{T \leq O\}}$, es decir, sólo se sabe si el tiempo de supervivencia ocurrió antes o después de O . Nótese que los datos de estado actual son parecidos a los datos con censura por la derecha, sin embargo en los datos censurados por la derecha existe la posibilidad de observar el tiempo T y en este caso no.

Por ejemplo, un estudio sobre la presencia de SIDA en personas infectadas con el VIH, el evento de interés es la presencia de la enfermedad. Al grupo de individuos se les hará un seguimiento médico al tiempo O y en ese momento se sabe si se ha presentado evidencia de la enfermedad o no.

- Caso II. En este caso, sólo se sabe que el tiempo de supervivencia T ha ocurrido antes, durante o después de un intervalo de tiempo. Este tipo de datos aparece en estudios médicos en los que el paciente es sometido a una revisión periódica. Sean O_1 y O_2 dos tiempos de observación, con $O_1 < O_2$ es decir, observación izquierda y derecha, respectivamente. Una observación de este tipo es el vector aleatorio de la forma $(O_1, O_2, \delta_1, \delta_2)$ con $\delta_1 = \mathbb{1}_{\{T \leq O_1\}}$ y $\delta_2 = \mathbb{1}_{\{O_1 < T \leq O_2\}}$, respectivamente.

Por ejemplo, si en el estudio sobre la presencia de SIDA se hace más de una revisión a cada individuo, los datos corresponden a este caso.

Un caso particular de este tipo de censura son los datos de la forma (O_1, O_2) donde el tiempo de supervivencia T cumple $O_1 < T < O_2$, este tipo de datos son **datos con tiempo de supervivencia doblemente censurado**, los cuales no debemos confundir con los datos doblemente censurados. Autores como De Gruttola y Lagakos (1989), Kim y otros (1993) y Gómez y Lagakos (1994) han propuesto modelos de supervivencia para datos con tiempo de supervivencia doblemente censurado.

Por otro lado, la censura por intervalo es una generalización de la censura por

la derecha y por la izquierda, ya que cuando el extremo inferior del intervalo es 0 y el superior C se tiene censura por la derecha y cuando el extremo inferior es C y el superior infinito, se tiene censura por la izquierda.

Es obvio que en un estudio se puedan encontrar combinaciones de las tres censuras: por la derecha, por la izquierda o de intervalo, y evidentemente, el tipo de censura influye en la construcción de la función de verosimilitud del modelo a partir de la cual se obtiene la estimación de los parámetros de los modelos de supervivencia como se verá más adelante.

1.1.2. Funciones Importantes para Datos de Supervivencia

En análisis de supervivencia hay dos funciones básicas y de interés, estas funciones son la *función de supervivencia* y la *función de riesgo*. Antes de definir dichas funciones, es necesario aclarar que, a lo largo de este trabajo, las letras en minúscula denotarán valores de variables aleatorias y las letras mayúsculas las variables aleatorias.

Sea t el tiempo hasta un evento específico, es decir, t es un valor de la variable aleatoria T no-negativa. Supóngase que la variable aleatoria T tiene una distribución de probabilidad P_T entonces la función de distribución viene dada por

$$F_T(t) = P_T(T < t) = \int_0^t f_T(u)du,$$

donde $f_T(\cdot)$ es la función de densidad.

La función más básica en análisis de supervivencia es la *función de supervivencia*, es decir, la probabilidad de que un individuo sobreviva más allá o igual al tiempo t y es definida como

$$S_T(t) = P_T(T \geq t),$$

la función de supervivencia es una función no-creciente tal que

$$S_T(0) = 1$$

y

$$S_T(\infty) = 0.$$

Si T es una variable aleatoria continua, la función de supervivencia es el complemento de la función de distribución acumulativa

$$S_T(t) = 1 - F_T(t)$$

esto es, la probabilidad de que el tiempo de supervivencia sea mayor que t . Como $f_T(\cdot)$ es la función de densidad de probabilidad se tiene

$$S_T(t) = P_T(T \geq t) = \int_t^{+\infty} f_T(u) du = F_T([t, +\infty]). \quad (1.1.1)$$

Por otro lado, otra función básica en análisis de supervivencia es la *función de riesgo* y es la probabilidad de que un individuo muera al tiempo t dado que ha sobrevivido a este tiempo. Representa el riesgo instantáneo de morir en el tiempo t y se define como

$$\lambda_T(t) = \lim_{\Delta t \rightarrow 0} \frac{P_T(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t}. \quad (1.1.2)$$

Por la probabilidad condicional se tiene que

$$\begin{aligned} P_T(t \leq T \leq t + \Delta t | T \geq t) &= \frac{P_T(t \leq T \leq t + \Delta t, t \leq T)}{P_T(t \leq T)} \\ &= \frac{P_T(t \leq T \leq t + \Delta t)}{P_T(t \leq T)}, \end{aligned}$$

y por otro lado, se sabe que

$$P_T(T \leq t + \Delta t) = \int_0^{t+\Delta t} f_T(u) du = \int_0^t f_T(u) du + \int_t^{t+\Delta t} f_T(u) du$$

entonces

$$F_T(t + \Delta t) = F_T(t) + P_T(t \leq T < t + \Delta t)$$

de donde

$$P_T(t \leq T < t + \Delta t) = F_T(t + \Delta t) - F_T(t).$$

Así, por (1.1.1) se tiene que

$$\begin{aligned}
 \lambda_T(t) &= \lim_{\Delta t \rightarrow 0} \left\{ \frac{F_T(t + \Delta t) - F_T(t)}{S_T(t)} \frac{1}{\Delta t} \right\} \\
 &= \lim_{\Delta t \rightarrow 0} \left\{ \frac{F_T(t + \Delta t) - F_T(t)}{\Delta t} \right\} \frac{1}{S_T(t)} \\
 &= \frac{f_T(t)}{S_T(t)}
 \end{aligned} \tag{1.1.3}$$

ya que $f_T(\cdot)$ es la primitiva de $F_T(\cdot)$.

A partir de (1.1.3) se pueden obtener algunas relaciones entre estas funciones que son muy utilizadas en análisis de supervivencia, por ejemplo, se tiene

$$\begin{aligned}
 \lambda_T(t) &= \frac{f_T(t)}{S_T(t)} \\
 &= \frac{f_T(t)}{1 - F_T(t)} \\
 &= -\frac{d}{dt} (\log (1 - F_T(t))) \\
 &= -\frac{d}{dt} (\log S_T(t)).
 \end{aligned} \tag{1.1.4}$$

De aquí, se obtiene la expresión

$$-\Lambda_T(t) = \log (S_T(t)) \tag{1.1.5}$$

con

$$\Lambda_T(t) = \int_0^t \lambda_T(u) du, \tag{1.1.6}$$

donde la función $\Lambda_T(\cdot)$ se llama *función de riesgo acumulativa*. Además, por la ecuación (1.1.3) y (1.1.1) se puede expresar la función de riesgo acumulativa como

$$\begin{aligned}
 \Lambda_T(t) &= \int_0^t \frac{f_T(u)}{S_T(u)} du \\
 &= \int_0^t \frac{F_T(du)}{F_T([u, +\infty])}
 \end{aligned}$$

por tanto, se puede considerar Λ_T como una medida

$$\Lambda_T(dt) = \frac{F_T(dt)}{F_T([0, +\infty])}, \quad (1.1.7)$$

donde dt es el intervalo $[t, t + dt]$.

1.1.3. Función de Verosimilitud

Una vez que se elige un modelo estadístico apropiado para un conjunto de datos de supervivencia, los parámetros de éste serán estimados maximizando la función de verosimilitud. Puesto que los datos de supervivencia están sujetos a mecanismos de censura, se debe poner especial cuidado en la construcción de la función de verosimilitud. A lo largo de esta tesis, se asume que los tiempos de supervivencia T y de censura C son independientes.

Se considera el caso con datos doblemente censurados, es decir, datos con censura por la derecha en el tiempo c^D y datos con censura por la izquierda en el tiempo c^I . Nótese que un dato correspondiente a un dato completo con tiempo de supervivencia t nos da información de la probabilidad de que el evento ocurra en este tiempo t aproximadamente igual a la función de densidad en este punto, esto contribuye con $f(t)$. Para un dato con censura por la derecha en el tiempo c^D se sabe que el tiempo de supervivencia es mayor que el tiempo c^D , entonces se toma la información de la función de supervivencia en este punto $S(c^D)$. Análogamente, para datos con censura por la izquierda en un tiempo c^I , se sabe que el evento ha ocurrido por tanto su contribución a la función de verosimilitud es el valor de la función acumulativa evaluada en el punto c^I , esto es $F(c^I) = 1 - S(c^I)$.

Por lo anterior, la función de verosimilitud debe cumplir

$$\mathcal{L} \propto \prod_{i \in \mathcal{C}} f_i(t_i) \prod_{i \in \mathcal{D}} S_i(c_i^D) \prod_{i \in \mathcal{I}} (1 - S_i(c_i^I)) \quad (1.1.8)$$

donde \mathcal{C} , \mathcal{D} e \mathcal{I} son el conjunto de datos completos, con censura por la derecha e

izquierda, respectivamente. Obsérvese que los subíndices i en las funciones f y S es debido a que cada individuo tiene diferente función de densidad.

La función de verosimilitud (1.1.8) es general. A continuación, se describirá la función de verosimilitud para datos con censura por la derecha, ya que es el mecanismo de censura más común y para la que presentamos, en las secciones restantes de este capítulo, uno de los modelos más conocidos en análisis de supervivencia, el modelo de Cox.

Se asume que se tiene una muestra independiente e idénticamente distribuida (iid) de observaciones de la forma (X_i, δ_i) con $i = 1, \dots, n$ con $X_i = \min(T_i, C_i)$ y $\delta_i = \mathbb{1}_{\{X_i = C_i\}}$, es decir, se tienen datos con censura por la derecha. Para esta muestra la ecuación (1.1.8) se puede expresar como

$$\mathcal{L}_n \propto \prod_{i=1}^n [f_i(x_i)]^{\delta_i} [S_i(x_i)]^{1-\delta_i} \quad (1.1.9)$$

debido a que δ_i nos indica a cual conjunto pertenece la observación i .

Para un modelo conocido de supervivencia se tendrá la forma específica de las funciones $f_i(\cdot)$ y $S_i(\cdot)$ con la que se obtendrá la función de verosimilitud particular para el correspondiente modelo.

Por otra parte, en los datos de estado actual sólo se sabe si el evento ha ocurrido o no al momento de hacer la observación en tiempo x . Si el tiempo de supervivencia ya ha ocurrido al tiempo x contribuye en la función de verosimilitud con el valor de la función acumulativa evaluada en el punto x , esto es con el valor $F(x) = 1 - S(x)$. Si por el contrario, el tiempo de supervivencia no ha ocurrido aún en el tiempo de observación x entonces se sabe que el tiempo de supervivencia es mayor que el tiempo x , por tanto influirá en la función de verosimilitud con el valor $S(x)$. Entonces, la función de verosimilitud viene dada

$$\mathcal{L} \propto \prod_{i \in \mathcal{B}} (1 - S_i(x_i)) \prod_{i \in \mathcal{A}} S_i(x_i) \quad (1.1.10)$$

donde \mathcal{B} y \mathcal{A} son los conjuntos de datos tales que el tiempo de supervivencia ha ocurrido antes o después al tiempo de la observación, respectivamente. Los subíndices de la función $S(\cdot)$ indican que la función de supervivencia es diferente para cada individuo.

Si se asume que se tiene una muestra iid de vectores (X_i, δ_i) con $i = 1, \dots, n$ donde X_i es el tiempo de observación y $\delta_i = \mathbb{1}_{\{T_i \leq X_i\}}$ donde T_i es el tiempo de supervivencia, es decir, datos de estado actual, entonces, la función (1.1.10) se puede expresar de la forma

$$\mathcal{L}_n \propto \prod_{i=1}^n (1 - S_i(x_i))^{\delta_i} (S_i(x_i))^{1-\delta_i}.$$

1.2. Modelo de Cox de Riesgos Proporcionales

Muchas veces, cuando se tienen datos de supervivencia, también se tiene información adicional de cada individuo. Por ejemplo, en el caso de los ensayos clínicos, generalmente se tienen datos de cada paciente como la edad, edad de diagnóstico de la enfermedad, sexo, valores de variables fisiológicas, si recibe otro u otros tratamientos, etc., todos estos factores pueden influir en la supervivencia del paciente, por lo cual tiene sentido incluir esta información en el modelo para el análisis de supervivencia, estos datos se incluyen como variables explicativas, llamadas *covariables*.

Una forma de utilizar estas covariables es modelando la función de riesgo $\lambda_T(\cdot)$, en cierto sentido, como en un modelo de regresión, con algunas diferencias particulares de acuerdo a los datos y análisis de supervivencia.

En este contexto, el modelo más usado es el modelo propuesto por Cox (1972), también llamado *modelo de riesgos proporcionales*. En esta sección se describe brevemente dicho modelo, así como las referencias básicas donde se han propuesto

la estimación de parámetros y se han establecido las propiedades asintóticas de estos, entre otras cosas.

Se considera la función de riesgo $\lambda(\cdot|\mathbf{z})$ en un tiempo t y para un individuo con vector de covariables \mathbf{z} , es decir $t \in \mathbb{R}^+$ y $\mathbf{z} \in \mathbb{R}^d$. El modelo de regresión de Cox (1972) propone que

$$\lambda(t|\mathbf{z}) = \lambda_0(t)\varphi(\boldsymbol{\beta}, \mathbf{z}) \quad (1.2.11)$$

donde $\lambda_0(\cdot)$ es una función de riesgo base y $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)$ es el vector de parámetros del modelo y la función $\varphi(\cdot, \cdot)$ es la función nexo de las variables explicativas. Una posibilidad es definir $\varphi(\boldsymbol{\beta}, \mathbf{z})$ como $\exp(\boldsymbol{\beta}'\mathbf{z})$ debido a que es deseable que la función $\varphi(\cdot, \cdot)$ sea no negativa. A este modelo se le conoce como *modelo de Cox* y viene dado por

$$\lambda(t|\mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z}). \quad (1.2.12)$$

Nótese que si se toma

$$\log \frac{\lambda(t|\mathbf{z})}{\lambda_0(t)} = \boldsymbol{\beta}'\mathbf{z}$$

se obtiene un modelo de regresión lineal tradicional.

En el mismo sentido que en regresión lineal, se pueden usar variables de tipo factor y transformarlas en vectores indicadores o construir modelos con interacción entre las covariables considerando el producto entre ellos. Más adelante se describirá el procedimiento para convertir una variable tipo factor en variables indicadoras, ya que será una técnica que se aplicará en los datos reales que se usan para ilustrar los modelos propuestos.

El modelo de Cox (1.2.12) se llama de riesgos proporcionales porque si se calcula la razón entre las funciones de riesgo para las covariables \mathbf{z}_i y \mathbf{z}_j se obtiene una constante, esto es

$$\rho(t|\boldsymbol{\beta}, \mathbf{z}_i, \mathbf{z}_j) = \frac{\lambda(t|\mathbf{z}_i)}{\lambda(t|\mathbf{z}_j)} = \frac{\lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z}_i)}{\lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z}_j)} = \exp(\boldsymbol{\beta}'(\mathbf{z}_i - \mathbf{z}_j)), \quad (1.2.13)$$

que se denomina el *riesgo relativo* de un individuo con covariable \mathbf{z}_i , comparado con un individuo con covariable \mathbf{z}_j .

Por otro lado, usando la ecuación (1.1.5) y el modelo (1.2.12) se obtiene que la función de supervivencia dada por

$$S(t|\mathbf{z}) = \exp(-\Lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z})) \quad (1.2.14)$$

debido a que

$$S(t|\mathbf{z}) = \exp(-\Lambda(t|\mathbf{z}))$$

y

$$\begin{aligned} \Lambda(t|\mathbf{z}) &= \int_0^t \lambda(u|\mathbf{z}) du \\ &= \int_0^t \lambda_0(u) \exp(\boldsymbol{\beta}'\mathbf{z}) du \\ &= \exp(\boldsymbol{\beta}'\mathbf{z}) \int_0^t \lambda_0(u) du \\ &= \exp(\boldsymbol{\beta}'\mathbf{z}) \Lambda_0(t) \end{aligned}$$

donde $\Lambda_0(t) = \int_0^t \lambda_0(u) du$. Entonces la función de distribución del modelo de Cox es

$$F(t|\mathbf{z}) = 1 - \exp(-\Lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z})). \quad (1.2.15)$$

Por último, cabe mencionar que también se suele hacer referencia como modelo de Cox, al modelo expresado por

$$\Lambda(t|\mathbf{z}) = \Lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z}).$$

Hasta ahora no se ha hecho ningún supuesto sobre la función de riesgo base $\lambda_0(\cdot)$. Se puede suponer la función $\lambda_0(\cdot)$ como una función paramétrica, como por ejemplo una función Weibull o Exponencial y el modelo sería un modelo totalmente paramétrico. Sin embargo, en el resto del capítulo se considera que $\lambda_0(\cdot)$ es una función arbitraria no paramétrica, lo cual da mayor flexibilidad al modelo. Por

tanto, el modelo (1.2.12) es considerado como un modelo semiparamétrico como en Cox (1972), Breslow (1974) y Gill (1984). Bajo este marco de referencia, la estimación del modelo de Cox (1.2.12) implica estimar el vector de parámetros β y la función de riesgo base acumulativa $\Lambda_0(\cdot)$ dado un conjunto de observaciones de datos de supervivencia. Las propiedades asintóticas de estos estimadores han sido estudiados por Tsiatis (1981), quien consideró que las covariables \mathbf{z} son realizaciones de un vector aleatorio. Posteriormente, Johnson y otros (1982) realizaron un estudio de estos estimadores para muestras pequeñas.

Como se dijo anteriormente, en los datos de supervivencia se pueden encontrar variables de tipo factor, por tanto, a continuación, se explica la metodología para transformar una covariable de tipo factor en un vector de covariables indicadoras, aunque esta metodología es la misma que suele usarse en análisis de regresión clásico.

Transformación de las covariables de tipo factor

Supóngase que la función de riesgo $\lambda(\cdot)$ depende de una covariable de tipo factor A con a niveles. Se puede expresar la covariable A como una combinación de $a - 1$ variables explicativas indicadoras Z_1, \dots, Z_{a-1} como se muestra en la siguiente tabla:

nivel	Z_1	Z_2	...	Z_{a-1}
1	0	0	...	0
2	1	0	...	0
3	0	1	...	0
\vdots				
a	0	0	...	1

es decir, se tienen $a - 1$ parámetros asociados al efecto del factor A con a niveles.

1.2.1. Función de Verosimilitud Parcial

Se considera un conjunto de datos de supervivencia donde puede existir censura por la derecha, esto es, se tiene una muestra iid de vectores $(X_i, \delta_i, \mathbf{Z}_i)$ donde $X_i = \min(T_i, C_i)$, $\delta_i = \mathbb{1}_{\{X_i=C_i\}}$, con T_i y C_i el respectivo tiempo de supervivencia y censura y δ_i la variable indicadora si lo que se observa es un tiempo de censura o de supervivencia, en este caso, también se observa el vector de covariables \mathbf{Z}_i para cada individuo i con $i = 1, \dots, n$.

Como se desconoce la función $\lambda_0(\cdot)$ no se puede determinar la función de densidad del modelo, por tanto no se puede utilizar la expresión (1.1.9) de la función de verosimilitud para la muestra considerada. Sin embargo, en Cox (1972) y Cox (1975) se propone el método de verosimilitud parcial para la estimación del vector de parámetros β . La metodología es la siguiente:

En la muestra de n individuos, supóngase que hay r tiempos de supervivencia diferentes y $n - r$ datos censurados. Sin pérdida de generalidad, se puede suponer que los tiempos de supervivencia están ordenados tal que

$$t_1 < t_2 < \dots < t_r.$$

Se denota por $R(t_j)$ el conjunto de individuos en riesgo de morir al tiempo t_j , este conjunto está formado por los individuos con vida y sin censura antes del tiempo t_j .

La construcción de la función de verosimilitud parcial se basa en el argumento de que la función de riesgo $\lambda_0(\cdot)$ es cero en los tiempos que no son tiempos de supervivencia y por tanto, $\lambda(\cdot)$ también es cero en estos tiempos. Entonces, los intervalos de tiempo entre t_i y t_{i+1} no aportan información sobre el valor del vector de parámetros del modelo (1.2.12).

Cox (1972) considera la probabilidad de que un individuo j muera al tiempo t_j condicionado a que t_j es un tiempo de supervivencia observado con vector de

covariables \mathbf{z}_j , esto es

$$P(\text{que un individuo con covariable } \mathbf{z}_j \text{ muera en } t_j | \text{ hay una muerte en } t_j)$$

usando el Teorema de Bayes se tiene que es igual a

$$\begin{aligned} & \frac{P(\text{que un individuo con covariable } \mathbf{z}_j \text{ muera en } t_j, \text{ hay una muerte en } t_j)}{P(\text{hay una muerte en } t_j)} \\ &= \frac{P(\text{que un individuo con covariable } \mathbf{z}_j \text{ muera en } t_j)}{P(\text{hay una muerte en } t_j)} \end{aligned} \quad (1.2.16)$$

Se asume que los tiempos de muerte son independientes unos de otros entonces

$$P(\text{hay una muerte en } t_j) = \sum_{\ell \in R(t_j)} P(\text{que el individuo } \ell \text{ muera en } t_j)$$

es decir, la probabilidad de una muerte en t_j es la suma de las probabilidades de que mueran los individuos del conjunto de riesgo. Entonces la expresión (1.2.16) es igual a

$$\frac{P(\text{que un individuo con covariable } \mathbf{z}_j \text{ muera en } t_j)}{\sum_{\ell \in R(t_j)} P(\text{que el individuo } \ell \text{ muera en } t_j)}. \quad (1.2.17)$$

Posteriormente, se considera la probabilidad de morir en el intervalo de tiempo $(t_j, t_j + \Delta t)$, por la expresión (1.2.17) se tiene

$$\frac{P(\text{que un individuo con covariable } \mathbf{z}_j \text{ muera en } (t_j, t_j + \Delta t))}{\sum_{\ell \in R(t_j)} P(\text{que el individuo } \ell \text{ muera en } (t_j, t_j + \Delta t))}$$

si se dividen ambos términos de la expresión anterior por Δt y se toma el límite cuando $\Delta t \rightarrow 0$ se tiene la correspondiente función de riesgo $\lambda(\cdot)$ por la definición de la ecuación (1.1.2). De donde se obtiene la expresión

$$\frac{\lambda(t_j | \mathbf{z}_j)}{\sum_{\ell \in R(t_j)} \lambda(t_j | \mathbf{z}_\ell)} \quad (1.2.18)$$

y por la expresión (1.2.12) del modelo de Cox se tiene

$$\begin{aligned} &= \frac{\lambda_0(t_j) \exp(\boldsymbol{\beta}' \mathbf{z}_j)}{\sum_{\ell \in R(t_j)} \lambda_0(t_j) \exp(\boldsymbol{\beta}' \mathbf{z}_\ell)} \\ &= \frac{\exp(\boldsymbol{\beta}' \mathbf{z}_j)}{\sum_{\ell \in R(t_j)} \exp(\boldsymbol{\beta}' \mathbf{z}_\ell)}. \end{aligned}$$

Finalmente, se obtiene la *función de verosimilitud parcial* tomando el producto de todas las probabilidades condicionales sobre r y viene dada por

$$\mathcal{L}_n(\boldsymbol{\beta}) = \prod_{j=1}^r \frac{\exp(\boldsymbol{\beta}' \mathbf{z}_j)}{\sum_{\ell \in R(t_j)} \exp(\boldsymbol{\beta}' \mathbf{z}_\ell)} = \prod_{i=1}^n \left(\frac{\exp(\boldsymbol{\beta}' \mathbf{z}_i)}{\sum_{\ell \in R(x_i)} \exp(\boldsymbol{\beta}' \mathbf{z}_\ell)} \right)^{\delta_i}. \quad (1.2.19)$$

Se llama función de verosimilitud parcial ya que no toma toda la información de los datos completos y censurados como la función (1.1.9).

Tomando logaritmo en (1.2.19), la función de log-verosimilitud parcial viene dada por

$$\begin{aligned} \ell_n(\boldsymbol{\beta}) &= \log \mathcal{L}_n(\boldsymbol{\beta}) \\ &= \log \prod_{i=1}^n \left(\frac{\exp(\boldsymbol{\beta}' \mathbf{z}_i)}{\sum_{\ell \in R(x_i)} \exp(\boldsymbol{\beta}' \mathbf{z}_\ell)} \right)^{\delta_i} \\ &= \sum_{i=1}^n \delta_i \left(\boldsymbol{\beta}' \mathbf{z}_i - \log \sum_{\ell \in R(x_i)} \exp(\boldsymbol{\beta}' \mathbf{z}_\ell) \right). \end{aligned} \quad (1.2.20)$$

Por tanto, los estimadores de máxima verosimilitud parcial de los parámetros del modelo serán

$$\hat{\boldsymbol{\beta}}_n = \arg \max_{\boldsymbol{\beta}} \ell_n(\boldsymbol{\beta}).$$

En la práctica, para obtener $\hat{\boldsymbol{\beta}}_n$, se usan métodos numéricos para encontrar el máximo de la función de log-verosimilitud parcial (1.2.20). Estos métodos numéricos suelen ser del tipo Newton–Raphson.

Cabe mencionar que las propiedades teóricas de este estimador como consistencia y eficiencia han sido establecidas por diversos autores, siendo Andersen y Gill (1982) uno de los más referenciados.

Hasta el momento, sólo se ha solucionado una parte del problema de estimación del modelo, ya que la función $\Lambda_0(\cdot)$ continúa siendo desconocida. En el siguiente apartado se mostrará como obtener el estimador $\hat{\Lambda}_0(\cdot)$.

1.2.2. Función de Verosimilitud Total

Otra metodología para estimar los parámetros del modelo de Cox (1.2.12) fue propuesta por Johansen (1983), considerando la función de verosimilitud total. A continuación se describe la metodología.

Se considera la función de verosimilitud (1.1.9) para datos con censura por la derecha, salvo constantes, dada por

$$\mathcal{L}_n(\boldsymbol{\beta}, \lambda_0) = \prod_{i=1}^n [f_i(x_i)]^{\delta_i} [S_i(x_i)]^{1-\delta_i}$$

por la ecuación (1.1.3) se tiene

$$\begin{aligned} \mathcal{L}_n(\boldsymbol{\beta}, \lambda_0) &= \prod_{i=1}^n [\lambda_i(x_i) S_i(x_i)]^{\delta_i} [S_i(x_i)]^{1-\delta_i} \\ &= \prod_{i=1}^n [\lambda_i(x_i)]^{\delta_i} S_i(x_i). \end{aligned} \tag{1.2.21}$$

Hasta el momento las funciones $\lambda_i(\cdot)$ y $S_i(\cdot)$ son funciones arbitrarias de supervivencia. Si se considera el modelo de Cox (1.2.12), la función de verosimilitud total debe estar en función de $\boldsymbol{\beta}$, $\lambda_0(\cdot)$ ya que son los parámetros desconocidos. Así, considerando la expresión (1.2.14) se puede escribir la expresión (1.2.21) como

$$\mathcal{L}_n(\boldsymbol{\beta}, \lambda_0) = \prod_{i=1}^n [\lambda_0(x_i) \exp(\boldsymbol{\beta}' \mathbf{z}_i)]^{\delta_i} \left(\exp \left(- \exp(\boldsymbol{\beta}' \mathbf{z}_i) \int_0^{x_i} \lambda_0(u) du \right) \right).$$

Si se define $Y_i(x) = \exp(\beta' \mathbf{z}_i) \mathbb{1}_{\{X_i \geq x\}}$ e $Y(x) = \sum_i Y_i(x)$ se puede escribir la verosimilitud total como

$$\begin{aligned} \mathcal{L}_n(\beta, \lambda_0) &= \prod_{i=1}^n [\lambda_0(x_i) \exp(\beta' \mathbf{z}_i)]^{\delta_i} \prod_{i=1}^n \left[\frac{\lambda_0(x_i) Y(x_i)}{\lambda_0(x_i) Y(x_i)} \right]^{\delta_i} \exp \left(- \exp(\beta' \mathbf{z}_i) \int_0^{x_i} \lambda_0(u) du \right) \\ &= \prod_{i=1}^n \left[\frac{\exp(\beta' \mathbf{z}_i)}{Y(x_i)} \right]^{\delta_i} \prod_{i=1}^n [\lambda_0(x_i) Y(x_i)]^{\delta_i} \exp \left(- \exp(\beta' \mathbf{z}_i) \int_0^{x_i} \lambda_0(u) du \right). \end{aligned} \quad (1.2.22)$$

Obsérvese que la verosimilitud parcial propuesta por Cox (1972) es el primer factor de (1.2.22). Johansen (1983) propone usar el segundo factor de (1.2.22) para estimar $\lambda_0(\cdot)$, ya que maximizando la log-verosimilitud parcial se obtiene $\hat{\beta}$. El segundo término de la verosimilitud total puede escribirse como

$$\prod_{i=1}^n [\lambda_0(x_i) Y(x_i)]^{\delta_i} \exp \left(- \int_0^\infty Y(u) \lambda_0(u) du \right). \quad (1.2.23)$$

Por otro lado, se supone que I_i es el intervalo de tiempo que contiene x_i y que la función $\lambda_0(\cdot)$ es constante a trozos de longitud ε y además

$$\lambda_0(u) = 0 \quad u \notin \bigcup_i I_i$$

entonces la expresión (1.2.23) es igual a

$$\prod_{i=1}^n \left[Y(x_i) \lambda_0(x_i) \exp \left(- \int_{I_i} Y(u) \lambda_0(x_i) du \right) \right]$$

tomando el logaritmo se tiene

$$\sum_{i=1}^n \log (Y(x_i) \lambda_0(x_i)) - \lambda_0(x_i) \int_{I_i} Y(u) du.$$

Como para cada x_i la función $\lambda_0(x_i)$ es una constante, se puede derivar y obtener que el mínimo debe cumplir

$$\frac{1}{\lambda_0(x_i)} = \int_{I_i} Y(u) du$$

por tanto

$$\hat{\lambda}_0(x_i) = \left[\int_{I_i} Y(u) du \right]^{-1}$$

y se puede aproximar por

$$\hat{\lambda}_0(x_i) \simeq [\varepsilon Y(x_i)]^{-1}.$$

Al ser

$$\hat{\lambda}_0(x_i) \xrightarrow{\varepsilon \rightarrow 0} \infty$$

entonces se puede estimar $\Lambda_0(\cdot)$ como

$$\hat{\Lambda}_0(A) = \sum_{x_i \in A} \left(\frac{1}{Y(x_i)} \right).$$

Este es el estimador de Breslow (1974) para $\Lambda_0(\cdot)$, que es uno de los más utilizados.

Sustituyendo $\hat{\Lambda}_0(\cdot)$ en (1.2.22) y simplificando se llega a una verosimilitud “profile” proporcional a la verosimilitud parcial dada en (1.2.19). En diversos artículos como Murphy y van der Vaart (2000) e Hirose (2010) se establecen las propiedades asintóticas de los estimadores del modelo semiparamétrico de Cox obtenidos a partir de la función de verosimilitud “profile”. Además, en Hirose (2010) establece la equivalencia entre los estimadores de máxima verosimilitud parcial y “profile”.

Por otro lado, un punto muy importante en análisis de supervivencia es poder estimar la función de supervivencia para un individuo bajo algún modelo. En la literatura diversos estimadores $\hat{S}(\cdot|\mathbf{z})$ de la función de supervivencia para el modelo de Cox han sido propuestos por Cox (1972), Kalbfleish y Prentice (1973), Breslow (1974) y Prentice y Gloeckler (1978), entre otros. En Andersen y otros (1996) se hace una revisión de los más utilizados, se establece su equivalencia asintótica y se comparan mediante un estudio de simulación para muestras pequeñas.

1.2.3. Simulación de Tiempos de Supervivencia

Con la finalidad de verificar las propiedades y técnicas de estimación e inferencia propuestas en análisis de datos de supervivencia, muchas veces es necesario generar tiempos de supervivencia provenientes de un modelo específico. En este caso, se describe el método propuesto por Bender y otros (2005) para generar tiempos de

supervivencia a partir de un modelo de Cox de riesgos proporcionales que es el utilizado en esta tesis. Recuérdesse que el tiempo de supervivencia en un modelo de Cox (1.2.12) está relacionado con las covariables mediante la función de riesgo $\lambda(\cdot)$.

Supóngase que t es una variable aleatoria con función de distribución $F(\cdot)$, entonces $\tilde{u} = F(t)$ sigue una distribución uniforme $\mathcal{U}(0, 1)$ y también $u = (1 - \tilde{u}) \sim \mathcal{U}(0, 1)$. Se asume que t es el tiempo de supervivencia proveniente de un modelo de Cox (1.2.12), entonces

$$\begin{aligned}
 u &= 1 - \tilde{u} \\
 &= 1 - F(t|\mathbf{z}) \\
 &= S(t|\mathbf{z}) \\
 &= \exp(-\Lambda(t)) \\
 &= \exp(-\Lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z})).
 \end{aligned} \tag{1.2.24}$$

Si $\Lambda_0(\cdot)$ es invertible, el tiempo t puede expresarse como

$$t = \Lambda_0^{-1}(-\exp(-\boldsymbol{\beta}'\mathbf{z}) \log u)$$

donde $u \sim \mathcal{U}(0, 1)$, debido a

$$\begin{aligned}
 u &= \exp(-\Lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z})) \\
 \Leftrightarrow \quad \log u &= -\Lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z}) \\
 \Leftrightarrow \quad -\frac{\log u}{\exp(\boldsymbol{\beta}'\mathbf{z})} &= \Lambda_0(t) \\
 \Leftrightarrow \quad t &= \Lambda_0^{-1}\left(-\frac{\log u}{\exp(\boldsymbol{\beta}'\mathbf{z})}\right).
 \end{aligned}$$

Como caso particular, se supone que la función de riesgo está dada por una distribución Weibull con parámetros positivos γ y ν , entonces la inversa de la función de riesgo acumulativa es

$$\Lambda_0^{-1}(x) = \left(\frac{x}{\gamma}\right)^{\frac{1}{\nu}}.$$

Así, el tiempo t generado bajo el modelo de Cox (1.2.12) con función de riesgo base una distribución Weibull(γ, ν) se puede expresar como

$$t = \left(\frac{-1}{\gamma} \frac{\log u}{\exp(\beta' \mathbf{z})} \right)^{\frac{1}{\nu}}. \quad (1.2.25)$$

Este procedimiento para generar tiempos de supervivencia es el utilizado en los estudios de simulación llevados a cabo en este trabajo.

1.3. Modelos Semiparamétricos

En esta sección, se mencionan las principales ideas sobre las técnicas usadas en modelos semiparamétricos, ya que esta metodología será la que se usará para demostrar algunas propiedades de los modelos que se proponen en este trabajo. La teoría aquí presentada se basa en Kosorok (2008), van der Vaart y Wellner (1996) y van der Vaart (1998), quienes usan la siguiente notación:

Para P una medida en un espacio medible $(\mathcal{X}, \mathcal{A})$ y $f : \mathcal{X} \rightarrow \mathbb{R}^d$ una función medible, $Pf = P[f]$ denota la integral $\int f dP$, ó equivalentemente la esperanza $E_P[f(X)]$ de la función f de una variable aleatoria X distribuida de acuerdo a P .

Supóngase que se tiene una muestra de tamaño n , que consiste en un conjunto de observaciones que provienen de una distribución común P . Un *modelo estadístico* es la colección \mathcal{P} de todos los posibles valores de P , esto es $\{P \in \mathcal{P}\}$ sobre el mismo espacio muestral. Estos modelos pueden ser de la forma

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$$

donde Θ es un espacio paramétrico.

Un *modelo* es llamado *semiparamétrico* si θ tiene uno o más parámetros finitos y tiene uno o más parámetros infinito-dimensionales, en este caso el modelo se puede expresar como

$$\theta = (\beta, \eta) \mapsto P_{\beta, \eta} \quad (1.3.26)$$

donde β y η son el parámetro finito e infinito dimensional, respectivamente.

Ejemplo 1.3.1. *Modelo de Cox.* Supóngase que se tienen observaciones de la forma (T, \mathbf{Z}) , donde T es el tiempo de supervivencia y \mathbf{Z} el vector de covariables. La distribución de \mathbf{Z} es desconocida y la función de riesgo dado \mathbf{Z} está dada por el modelo de Cox (1.2.12)

$$\lambda(t|\mathbf{z}) = \lambda_0(t) \exp(\beta' \mathbf{z})$$

con $\lambda_0(\cdot)$ desconocida, así el vector del modelo semiparamétrico (1.3.26) se puede expresar como (β, λ_0) con $\beta \in \mathbb{R}^d$ y $\lambda_0(\cdot)$ el parámetro infinito dimensional.

En modelos semiparamétricos, no es posible utilizar la función de log-verosimilitud total $\ell_n(\beta, \eta)$ sobre la muestra iid \mathbf{x}_i , con $i = 1, \dots, n$ de observaciones, para estimar ambos parámetros a la vez, una alternativa para estimar dichos parámetros es usar la verosimilitud “profile” para β definida como

$$p\ell_n(\beta) = \sup_{\eta} \ell_n(\beta, \eta)$$

donde $\ell_n(\cdot, \cdot)$ denota el producto de las verosimilitudes $\ell(\cdot, \cdot)(\mathbf{x}_i)$ con \mathbf{x}_i el i -ésimo elemento de la muestra para $i = 1, \dots, n$. La verosimilitud “profile” se puede considerar como una verosimilitud total en dos pasos: ya que en el par $(\hat{\beta}_n, \hat{\eta}_n)$ que maximiza $\ell_n(\beta, \eta)$, la primer componente $\hat{\beta}_n$ maximiza también la verosimilitud “profile” $\beta \mapsto p\ell_n(\beta)$.

La justificación de usar la verosimilitud “profile” en modelos semiparamétricos es que el problema infinito dimensional se reduce a un problema finito dimensional, para ello se definirá un *submodelo menos favorable* en dimensión finita, de la misma dimensión de β . Una vez hecho esto se puede usar un desarrollo en serie de Taylor para controlar los términos de error.

En general, estimar los parámetros de un modelo es encontrar el valor $\psi(P)$ para alguna función $\psi : \mathcal{P} \rightarrow D$, donde D es el dominio de los parámetros.

Estimar los parámetros $\psi(P)$ de un modelo es más difícil que estimar los parámetros $\psi(P_\theta)$ de un submodelo paramétrico $P_\theta \in \mathcal{P}_\theta \subset \mathcal{P}$ con

$$\mathcal{P}_\theta = \{P_\theta : \theta \in \Theta\} \subset \mathcal{P}.$$

Supóngase que $\hat{\theta}_n$ es un estimador del parámetro 1-dimensional del modelo P_θ , se dice que $\hat{\theta}_n$ es *eficiente* si el límite de la varianza V de $\sqrt{n}(\hat{\theta}_n - \psi(P_\theta))$ es lo más pequeña posible. Si se tiene el caso d -dimensional, la inversa de la matriz de covarianza V es la *matriz de información de Fisher*.

Si existe un submodelo \mathcal{P}_0 tal que la información del estimador $\hat{\theta}_n$ es el mínimo de la información de todos los submodelos se llama *submodelo menos favorable*.

La mayoría de veces, es suficiente considerar submodelos \mathcal{P}_0 d -dimensionales que pasan a través de la verdadera distribución P y que son diferenciables en P . Entonces se toma la función

$$\mathbf{t} \mapsto P_{\mathbf{t}} \in \mathcal{P}$$

tal que en $\mathbf{t} = \mathbf{t}_0$ pasa a través de la verdadera distribución P y que para una función “score” $g(\mathbf{x}) = \frac{\partial}{\partial \mathbf{t}} \log p_{\mathbf{t}}(\mathbf{x})|_{\mathbf{t}=\mathbf{t}_0}$, con $p_{\mathbf{t}}(\cdot)$ la densidad de $P_{\mathbf{t}}$, se tiene

$$P_0 \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \log p(\mathbf{x}) - g(\mathbf{x}) \right\|^2 \rightarrow 0 \text{ cuando } \mathbf{t} \rightarrow \mathbf{t}_0$$

donde $p(\cdot)$ es la densidad de P . La colección de funciones “score” es llamada *conjunto tangente* del modelo \mathcal{P} en P y se denota por $\dot{\mathcal{P}}_P$. Además, como $\dot{\mathcal{P}}_P \subset L_2(P)$, se tiene que $\dot{\mathcal{P}}_P$ es un *espacio tangente*.

Por otro lado, se consideran submodelos $\mathbf{t} \mapsto P_{\mathbf{t}}$ tal que $\mathbf{t} \mapsto \psi(P_{\mathbf{t}})$ es diferenciable en $\mathbf{t} = \mathbf{t}_0$. Más precisamente, se dice que $\psi : \mathcal{P} \rightarrow D$ diferenciable en P relativo a un conjunto tangente $\dot{\mathcal{P}}_P$, si existe una función lineal continua $\dot{\psi}_P(P) \mapsto D$ tal que para el submodelo $\mathbf{t} \mapsto P_{\mathbf{t}}$ con función “score” $g(\cdot) \in \mathcal{P}_P$ se tiene

$$\left. \frac{\partial \psi(P_{\mathbf{t}})}{\partial \mathbf{t}} \right|_{\mathbf{t}=\mathbf{t}_0} \rightarrow \dot{\psi}_P g.$$

La función $\dot{\psi}_P$ se puede escribir como producto interior entre un vector fijo y una función medible $\tilde{\psi}_P : X \mapsto \mathbb{R}^d$

$$\dot{\psi}_P g = \langle \tilde{\psi}_P, g \rangle_P = \int \tilde{\psi}_P g dP. \quad (1.3.27)$$

Además, si la función $\tilde{\psi}_P$ está en la clausura del espacio lineal generado por el conjunto tangente, es llamada *función “score” eficiente*.

Si se considera un modelo semiparamétrico

$$\{P_{\beta, \eta} : \beta \in \mathcal{B}, \eta \in H\}$$

con \mathcal{B} un subconjunto abierto de \mathbb{R}^d y H un subconjunto arbitrario de dimensión infinita, entonces un funcional de interés es $\psi(P_{\beta, \eta}) = \beta$, en este caso, se usan como submodelos las funciones de la forma

$$\mathbf{t} \mapsto P_{\mathbf{t}, \eta_{\mathbf{t}}} \quad (1.3.28)$$

para curvas dadas $\mathbf{t} \mapsto \eta_{\mathbf{t}}(\beta, \eta)$ en el conjunto de parámetros H . La función “score” para tales submodelos (si existen) pueden expresarse como la suma de las derivadas parciales con respecto a β y η .

La función “score” para β es definida como la derivada parcial con respecto a β de la función de log-verosimilitud $\ell(\beta, \eta)$ esto es

$$\frac{\partial}{\partial \beta} \ell(\beta, \eta)(\mathbf{x}) = \dot{\ell}(\beta, \eta)(\mathbf{x}).$$

Por otro lado, supóngase que la función “score” para η tiene la forma

$$\left. \frac{\partial}{\partial \mathbf{t}} \ell(\beta, \eta_{\mathbf{t}})(\mathbf{x}) \right|_{\mathbf{t}=\mathbf{t}_0} = A_{\beta, \eta} h(\mathbf{x})$$

donde h es una dirección en la cual $\eta_{\mathbf{t}} \in H$ se aproxima a η corriendo sobre índices del conjunto H .

Así, tomando el submodelo (1.3.28), se tiene que la función de log-verosimilitud depende del parámetro \mathbf{t} , por tanto se puede escribir como $\ell(\mathbf{t}, \boldsymbol{\beta}, \eta)(\mathbf{x})$ y se obtiene

$$\begin{aligned} \dot{\ell}(\mathbf{t}, \boldsymbol{\beta}, \eta)(\mathbf{x}) \Big|_{\mathbf{t}=\mathbf{t}_0} &= \frac{\partial}{\partial \mathbf{t}} \log p_{\mathbf{t}, \eta_{\mathbf{t}}}(\mathbf{x}) \Big|_{\mathbf{t}=\mathbf{t}_0} \\ &= \dot{\ell}(\boldsymbol{\beta}, \eta)(\mathbf{x}) - A_{\boldsymbol{\beta}, \eta} h(\mathbf{x}). \end{aligned} \quad (1.3.29)$$

Por la expresión (1.3.27) se tiene

$$\frac{\partial}{\partial \mathbf{t}} \psi(P_{\mathbf{t}, \eta_{\mathbf{t}}}) \Big|_{\mathbf{t}=\mathbf{t}_0} = \langle \tilde{\psi}_{\boldsymbol{\beta}, \eta}, \dot{\ell}(\boldsymbol{\beta}, \eta) - A_{\boldsymbol{\beta}, \eta} h \rangle_{P_{\boldsymbol{\beta}, \eta}}$$

si la función ψ es diferenciable en el modelo.

Además, si $\tilde{\psi}_{\boldsymbol{\beta}, \eta}$ está en el conjunto tangente $\dot{\mathcal{P}}_{P_{\boldsymbol{\beta}, \eta}}$ y la función “score” $\dot{\ell}(\boldsymbol{\beta}, \eta) = 0$ se tiene

$$0 = \langle \tilde{\psi}_{\boldsymbol{\beta}, \eta}, -A_{\boldsymbol{\beta}, \eta} h \rangle_{P_{\boldsymbol{\beta}, \eta}},$$

entonces, se define $\Pi_{\boldsymbol{\beta}, \eta}$ como la proyección ortogonal sobre la clausura del espacio lineal generado por $\dot{\mathcal{P}}_{P_{\boldsymbol{\beta}, \eta}}$ en $L_2(P_{\boldsymbol{\beta}, \eta})$. Así, la función

$$\tilde{\ell}(\boldsymbol{\beta}, \eta) = \dot{\ell}(\boldsymbol{\beta}, \eta) - \Pi_{\boldsymbol{\beta}, \eta} \dot{\ell}_{\boldsymbol{\beta}, \eta} \quad (1.3.30)$$

es la *función “score” eficiente* para $\boldsymbol{\beta}$ y la matriz $\tilde{\mathbf{I}}_{\boldsymbol{\beta}, \eta} = P_{\boldsymbol{\beta}, \eta} \tilde{\ell}_{\boldsymbol{\beta}, \eta} \tilde{\ell}'_{\boldsymbol{\beta}, \eta}$ es la matriz de información eficiente de Fisher.

Ahora, por otro lado, también se usará teoría de procesos empíricos, por tanto, se describirán algunos conceptos básicos.

La *medida empírica* \mathbb{P}_n de una muestra aleatoria X_1, \dots, X_n en un espacio medible $(\mathcal{X}, \mathcal{A})$ es la medida discreta

$$\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$

donde δ_{X_i} es la medida de Dirac de la observación X_i .

Dada una colección \mathcal{F} de funciones medibles $f : \mathcal{X} \rightarrow \mathbb{R}$, la medida empírica induce una función de \mathcal{F} a \mathbb{R} dada por

$$f \mapsto \mathbb{P}_n f,$$

donde se aplica la medida empírica discreta uniforme \mathbb{P}_n a una muestra X_1, \dots, X_n , de la siguiente forma

$$\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^n f(X_i).$$

Sea P la distribución de X_i , la versión centrada y escalada de una función f es un *proceso empírico* \mathbb{G}_n dado por

$$f \mapsto \mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - P)f = \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(X_i) - Pf)$$

donde $Qf = \int f dQ$. Frecuentemente, se denota $\mathbb{G}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\delta_{X_i} - P)$.

Además, para una función uniforme f en \mathcal{F} dada, por la ley de los grandes números y el Teorema central del límite se tiene que

$$\mathbb{P}_n f \xrightarrow{c.s.} Pf$$

y

$$\mathbb{G}_n f \rightsquigarrow \mathcal{N}(0, P(f - Pf)^2)$$

si Pf existe y $Pf^2 < \infty$, respectivamente. Por otro lado, si se tiene

$$\|\mathbb{P}_n - P\|_{\mathcal{F}} \rightarrow 0$$

donde $\|\mathbb{Q}\|_{\mathcal{F}} = \sup\{|\mathbb{Q}f| : f \in \mathcal{F}\}$ y la convergencia es en probabilidad exterior, la familia \mathcal{F} se llama *clase Glivenko–Cantelli*.

Además, si

$$\sup_{f \in \mathcal{F}} |f(x) - P| < \infty \text{ para cada } x$$

entonces, el proceso empírico $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ se puede ver como una función en $\ell^\infty(\mathcal{F})$, por tanto

$$\mathbb{G}_n \sqrt{n}(\mathbb{P}_n - P) \rightsquigarrow \mathbb{G} \text{ en } \ell^\infty(\mathcal{F})$$

donde \mathbb{G} es el límite de la cola de un elemento Borel medible en $\ell^\infty(\mathcal{F})$. A la clase \mathcal{F} se le llama *clase Donsker*.

En van der Vaart y Wellner (1996) y van der Vaart (1998) se pueden ver los teoremas de permanencia de la clase Donsker, de donde se pueden obtener las siguientes clases Donsker:

Ejemplo 1.3.2. Si \mathcal{F} es la colección de funciones indicadoras de la forma $f_t = \mathbb{1}_{(-\infty, t]}$ con $t \in \mathbb{R}$, entonces \mathcal{F} es una clase Donsker.

Ejemplo 1.3.3. El conjunto de funciones uniformemente acotadas y de variación uniformemente acotada es una clase Donsker.

Ejemplo 1.3.4. La clase de funciones cuyas derivadas de orden k existe y son uniformemente acotadas por constantes M_k es una clase Donsker.

Ejemplo 1.3.5. Si \mathcal{F} es una clase Donsker tal que $\|P\|_{\mathcal{F}} < \infty$ y sea ϕ una función Lipschitz, entonces la clase de funciones $\phi(f)$ es una clase Donsker.

Ejemplo 1.3.6. Si \mathcal{F} y \mathcal{G} son clases Donsker y $\|P\|_{\mathcal{F} \cup \mathcal{G}} < \infty$, entonces $\mathcal{F} \wedge \mathcal{G}$, $\mathcal{F} \vee \mathcal{G}$ y $\mathcal{F} + \mathcal{G}$ son clases Donsker.

Ejemplo 1.3.7. Si las clases \mathcal{F} y \mathcal{G} son Donsker y uniformemente acotadas, entonces $\{fg : f \in \mathcal{F} \text{ y } g \in \mathcal{G}\}$ es una clase Donsker.

Ejemplo 1.3.8. Si \mathcal{F} es Donsker con $\|P\|_{\mathcal{F}} < \infty$ y $f > 0$ para cada $f \in \mathcal{F}$, entonces la clase $\{1/f : f \in \mathcal{F}\}$ es Donsker.

Ejemplo 1.3.9. Si la clase \mathcal{F} es Donsker, entonces es Glivenko–Cantelli.

Finalmente, para demostrar la normalidad asintótica de los estimadores de los modelos semiparamétricos se usará el Teorema 1 demostrado en Murphy y van der Vaart (2000).

Supóngase la existencia de un *submodelo menos favorable* $\mathbf{t} \mapsto P_{\mathbf{t}, \eta_{\mathbf{t}}}$, es decir que para cada parámetro $(\boldsymbol{\beta}, \eta)$ existe una función $\mathbf{t} \mapsto \eta_{\mathbf{t}}(\boldsymbol{\beta}, \eta)$ tal que la función $\mathbf{t} \mapsto \ell(\mathbf{t}, \boldsymbol{\beta}, \eta)(\mathbf{x})$ está definida por

$$\ell(\mathbf{t}, \boldsymbol{\beta}, \eta)(\mathbf{x}) = \log p_{\mathbf{t}, \eta_{\mathbf{t}}}(\mathbf{x})$$

que es dos veces continuamente diferenciable para todo \mathbf{x} , donde la primera y segunda derivada son denotadas por

$$\dot{\ell}(\mathbf{t}, \boldsymbol{\beta}, \eta)(\mathbf{x}) \text{ y } \ddot{\ell}(\mathbf{t}, \boldsymbol{\beta}, \eta)(\mathbf{x}),$$

respectivamente. El submodelo con parámetros $(\mathbf{t}, \eta_{\mathbf{t}})$ debe pasar por $(\boldsymbol{\beta}, \eta)$ en $\mathbf{t} = \boldsymbol{\beta}$, es decir

$$\eta_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \eta) = \eta \text{ para cada } (\boldsymbol{\beta}, \eta). \quad (1.3.31)$$

Además, debe cumplir que la función “score” para el parámetro \mathbf{t} del modelo con verosimilitud $\ell(\mathbf{t}, \boldsymbol{\beta}_0, \eta_0)$ en $\mathbf{t} = \boldsymbol{\beta}_0$ es la función “score” eficiente para $\boldsymbol{\beta}$, para los verdaderos parámetros $\boldsymbol{\beta}_0$ y η_0 , esto es

$$\dot{\ell}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0, \eta_0) = \tilde{\ell}_{\boldsymbol{\beta}_0, \eta_0}. \quad (1.3.32)$$

Por otro lado, para cada $\boldsymbol{\beta}$ fijo, sea $\hat{\eta}_{\boldsymbol{\beta}}$ supremo de la función de verosimilitud “profile”

$$p\ell_n(\boldsymbol{\beta}) = \ell_n(\boldsymbol{\beta}, \hat{\eta}_{\boldsymbol{\beta}})$$

se asume que para cada $\tilde{\boldsymbol{\beta}}_n \xrightarrow{P} \boldsymbol{\beta}_0$ se tiene

$$\hat{\eta}_{\tilde{\boldsymbol{\beta}}_n} \xrightarrow{P} \eta \quad (1.3.33)$$

y

$$P_0 \dot{\ell}(\boldsymbol{\beta}_0, \tilde{\boldsymbol{\beta}}_n, \hat{\eta}_{\tilde{\boldsymbol{\beta}}_n}) = o_p \left(\|\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| + n^{\frac{1}{2}} \right) \quad (1.3.34)$$

Teorema 1.3.10. *Las condiciones (1.3.31) - (1.3.34) se satisfacen y supóngase que las funciones*

$$(\mathbf{t}, \boldsymbol{\beta}, \eta) \mapsto \dot{\ell}(\mathbf{t}, \boldsymbol{\beta}, \eta)(\mathbf{x})$$

y

$$(\mathbf{t}, \boldsymbol{\beta}, \eta) \mapsto \ddot{\ell}(\mathbf{t}, \boldsymbol{\beta}, \eta)(\mathbf{x})$$

son continuas en $(\beta_0, \beta_0, \eta_0)$ para P_0 para casi todo \mathbf{x} (ó en medida). Además supóngase que existe un entorno V de $(\beta_0, \beta_0, \eta_0)$ tal que la clase de funciones

$$\{\dot{\ell}(\mathbf{t}, \beta, \eta) : (\mathbf{t}, \beta, \eta) \in V\}$$

es Donker con función envolvente cuadrado integrable y

$$\{\ddot{\ell}(\mathbf{t}, \beta, \eta) : (\mathbf{t}, \beta, \eta) \in V\}$$

es Glivenko–Cantelli con función envolvente integrable.

Entonces

$$\begin{aligned} \log p\ell_n(\tilde{\beta}_n) &= \log p\ell_n(\beta_0) + (\tilde{\beta}_n - \beta_0)' \sum_{i=1}^n \tilde{\ell}_0(\mathbf{x}_i) - \frac{1}{2}n(\tilde{\beta}_n - \beta_0)' \tilde{I}_0(\tilde{\beta}_n - \beta_0) \\ &\quad + o_{P_0}(\sqrt{n}\|\tilde{\beta}_n - \beta_0\| + 1)^2 \end{aligned} \quad (1.3.35)$$

con $\tilde{\ell}_0$ la función “score” eficiente e \tilde{I}_0 es su matriz de información eficiente de Fisher.

Demostración. La demostración se puede ver en Murphy y van der Vaart (2000). □

Corolario 1.3.11. Si se tiene la expresión (1.3.35), y además \tilde{I}_0 es invertible y $\tilde{\beta}_n$ es consistente, entonces el estimador $\tilde{\beta}_n$ es asintóticamente normal con media $\mathbf{0}$ y matriz de covarianza \tilde{I}_0^{-1} .

1.4. Organización de los otros capítulos

El objetivo de este trabajo es proponer modelos de regresión para análisis de datos de supervivencia. En particular, se estudian dos modelos de regresión. El primero, es una alternativa al modelo de Cox para datos con censura por la derecha, para ello se propone una familia de modelos generalizados de riesgos proporcionales

basada en la función de distribución logística generalizada del Tipo-I. De esta familia se estudian las versiones paramétrica y semiparamétrica, en ambos casos se proponen los estimadores del modelo y se estudian sus propiedades asintóticas. Con la finalidad de verificar dichas propiedades asintóticas para muestras pequeñas o moderadas, se presentan los resultados de un estudio de simulación. Además, se ajustan estos nuevos modelos paramétricos y semiparamétricos a datos reales de supervivencia.

El segundo modelo estudiado es el modelo de Cox tradicional pero en vez de considerar datos de supervivencia con censura por la derecha se consideran datos doblemente censurados. Para este tipo de datos se supone un submodelo latente y en base a éste se proponen los estimadores del modelo de Cox, de los cuales se estudian las propiedades asintóticas. Dichas propiedades asintóticas son verificadas empíricamente para muestras pequeñas o moderadas mediante un estudio de simulación. También, se ajusta el modelo de Cox a un conjunto de datos reales doblemente censurados.

Finalmente, para verificar los efectos de las covariables en el modelo, se proponen estadísticos basados en medidas de divergencia. Se estudia la distribución asintótica de dichos estadísticos con la finalidad de construir contrastes de hipótesis. También se obtiene la expresión de la potencia asintótica para dichos estadísticos. Por último, se hace un estudio de simulación con la finalidad de verificar y comparar el comportamiento de los nuevos estadísticos entre sí y con estadísticos clásicos como el de Wald y cociente de verosimilitud.

Este trabajo consta de 5 capítulos. En el Capítulo 2, se define y estudia la familia de modelos de regresión generalizados de riesgos proporcionales para datos de supervivencia con censura por la derecha. La función de riesgo de esta familia de modelos utiliza como función nexo la función de distribución logística generalizada del Tipo-I. Después de definir el modelo, se describe la interpretación del modelo.

Posteriormente se proponen estimadores para la versión paramétrica y semiparamétrica, basándose en estimadores de máxima verosimilitud total para la versión paramétrica y en estimadores de máxima verosimilitud parcial para la versión semiparamétrica. Para estudiar las propiedades asintóticas de los estimadores del modelo semiparamétrico se considera la verosimilitud “profile”, no sin verificar que los estimadores de máxima verosimilitud parcial y máxima verosimilitud “profile” son equivalentes. También, se propone un estimador de la función de riesgo base, el cual sirve para definir estimadores de la función de supervivencia. En este capítulo se lleva a cabo un estudio de simulación tanto para verificar el comportamiento de los estimadores en muestras pequeñas, tanto como para comparar diferentes estimadores de la función de supervivencia del modelo. Cabe mencionar que antes de mostrar los resultados de simulación, se describe la metodología que se sigue para generar los datos de supervivencia del modelo generalizado de riesgos proporcionales. También, se ajustan dichos modelos paramétrico y semiparamétrico a datos reales de supervivencia.

En el Capítulo 3, se consideran datos doblemente censurados. Es necesario mencionar que los trabajos que consideran este tipo de datos se han centrado en estimar la función de supervivencia para datos de supervivencia sin covariables y no han propuesto modelos de regresión. En este trabajo se define un submodelo latente para los datos de supervivencia. Una vez hecho esto, se construye un estimador de la función de riesgo base del modelo semiparamétrico de regresión de Cox. Posteriormente, usando este estimador se propone un estimador de máxima verosimilitud de los parámetros del modelo. Luego, se establecen las propiedades asintóticas de los estimadores utilizando la teoría de modelos semiparamétricos. Posteriormente, se analiza el comportamiento de los estimadores para muestras finitas mediante un estudio de simulación. Finalmente, se ajusta el modelo de Cox a datos reales de supervivencia doblemente censurados.

En el Capítulo 4, se consideran los modelos estudiados en los Capítulos 2 y 3 y para ambos se proponen estadísticos basados en medidas de divergencias como las divergencias de Kullback–Leibler, de Jeffreys y de Rényi para contrastar hipótesis. Se muestra la distribución asintótica de los estadísticos propuestos, además de su potencia asintótica. Finalmente, como a pesar de saber que los estadísticos propuestos asintóticamente tienen el mismo comportamiento, no se sabe su comportamiento para muestras pequeñas, se hace un estudio de simulación para compararlos entre sí y con algunos estadísticos clásicos como el de Wald y el cociente de verosimilitud.

Finalmente, en el Capítulo 5 se presentan conclusiones y también se señalan algunos tópicos para investigaciones futuras. Además, el trabajo consta de tres apéndices. En el primer apéndice, se presentan las derivadas de las funciones de verosimilitud de los modelos propuestos, ya que éstas son utilizadas para maximizar numéricamente dichas funciones, con la finalidad de encontrar los estimadores del modelo. En el segundo apéndice se ilustra el método de integración por partes como es utilizado en este trabajo. Por último, en el tercer apéndice, se presentan datos de supervivencia de cáncer de laringe con censura por la derecha, utilizados en este trabajo para ilustrar el uso de los modelos propuestos.

Capítulo 2

Modelo Generalizado para Datos con Censura por la Derecha

2.1. Introducción

Los modelos de regresión para datos de supervivencia tradicionalmente se basan en el modelo de riesgos proporcionales de Cox (1972) dado por el modelo (1.2.12). Algunos trabajos recientes, han extendido los modelos de regresión de tal modo que el modelo de Cox sea un caso particular.

Por ejemplo, Etezadi y Ciampi (1987) proponen un modelo de la forma

$$\lambda(t|\boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{z}) = \exp(\boldsymbol{\alpha}'\mathbf{z})\lambda_0(t \exp(\boldsymbol{\beta}'\mathbf{z}))$$

donde $\boldsymbol{\alpha}$ y $\boldsymbol{\beta}$ son vectores de parámetros de regresión. Nótese que si $\boldsymbol{\beta} = \mathbf{0}$ se tiene el modelo de Cox y que si $\boldsymbol{\alpha} = \boldsymbol{\beta}$ se tiene el modelo de tiempo de fallo acelerado (ver Kalbfleish y Prentice (1980)). El modelo de tiempo de fallo acelerado es un modelo ampliamente utilizado en análisis de supervivencia. Los autores muestran que con este modelo se obtiene un mejor ajuste que con el modelo de Cox y que con el modelo de tiempo de fallo acelerado. Sin embargo, la función de riesgo base debe ser aproximada por “splines”.

Otra familia de modelos fue propuesta por Younes y Lachin (1997) y por Royston y Parmar (2002). Esta familia de modelos incluyen los modelos de riesgos proporcionales como caso particular. Se basan en la transformación de la función de supervivencia por una función nexa $g(\cdot)$ de tal modo que

$$g(S(t|\boldsymbol{\beta}, \mathbf{z})) = g(S_0(t)) + \boldsymbol{\beta}'\mathbf{z}.$$

El método de estimar la función de riesgo es por medio de B-“splines” o “splines” cúbicos.

Una alternativa al modelo de Cox es modelar la función de riesgo como

$$\lambda(t|\boldsymbol{\beta}, \mathbf{z}) = \lambda_0(t) \frac{\exp(\boldsymbol{\beta}'\mathbf{z})}{1 + \exp(\boldsymbol{\beta}'\mathbf{z})}$$

esto es, el efecto de las covariables se modela en escala logarítmica. El modelo anterior no ha sido estudiado ampliamente. Sin embargo, en MacKenzie (1996) y MacKenzie (1997) se considera un modelo logit con función de riesgo base constante y función de riesgo no proporcional de la forma

$$\lambda(t|\alpha, \boldsymbol{\beta}, \mathbf{z}) = \lambda_0 \frac{\exp(t\alpha + \boldsymbol{\beta}'\mathbf{z})}{1 + \exp(t\alpha + \boldsymbol{\beta}'\mathbf{z})}.$$

En este capítulo, se introduce una familia flexible de modelos de riesgos proporcionales para análisis de supervivencia generalizando la familia de los artículos MacKenzie (1996) y MacKenzie (1997). Se muestran las versiones paramétrica y semiparamétrica de la familia de modelos. Se consideran datos con censura por la derecha, además para la estimación de los parámetros del modelo, se proponen los métodos de máxima verosimilitud y máxima verosimilitud parcial, respectivamente para los casos paramétrico y semiparamétrico y se estudiarán las propiedades asintóticas de estos estimadores. Además, se proponen diferentes estimadores para la respectiva función de supervivencia, que se ilustrarán tanto con datos reales como con estudios de simulación.

2.2. Generalización del Modelo de Riesgos Proporcionales

La forma usual del modelo de Cox es considerar la función $\varphi(\cdot, \cdot)$ del modelo (1.2.11) como una función exponencial. En este capítulo se propone reemplazar la función $\varphi(\cdot, \cdot)$ del modelo (1.2.11) por una función de distribución logística generalizada del Tipo-I de la forma

$$F(y) = \left(\frac{1}{1 + \exp(-y)} \right)^a$$

con $-\infty < y < +\infty$ y $a > 0$. (Ver Balakrishnan (1992)).

Así, se define una familia de modelos de riesgos proporcionales cuya función de riesgo para un tiempo $t \in \mathbb{R}^+$ viene dada por

$$\lambda(t|\boldsymbol{\beta}, a, \mathbf{z}) = \lambda_0(t)K(\boldsymbol{\beta}, a, \mathbf{z}) \quad (2.2.1)$$

donde la función de riesgo base es la función $\lambda_0(\cdot)$ y la función nexo del modelo es

$$K(\boldsymbol{\beta}, a, \mathbf{z}) = \left(\frac{1}{1 + \exp(-\boldsymbol{\beta}'\mathbf{z})} \right)^a, \quad a > 0, \quad (2.2.2)$$

que depende del vector de covariables \mathbf{z} en \mathbb{R}^d , del vector de parámetros $\boldsymbol{\beta}$ en \mathbb{R}^d y del parámetro positivo a .

Para verificar la proporcionalidad del modelo, se considera el riesgo relativo ρ para las covariables \mathbf{z}_i y \mathbf{z}_j , que viene dado por:

$$\begin{aligned} \rho(t|\boldsymbol{\beta}, a, \mathbf{z}_i, \mathbf{z}_j) &= \frac{\lambda(t|\boldsymbol{\beta}, a, \mathbf{z}_i)}{\lambda(t|\boldsymbol{\beta}, a, \mathbf{z}_j)} \\ &= \frac{\left(\frac{\lambda_0(t)}{1 + \exp(-\boldsymbol{\beta}'\mathbf{z}_i)} \right)^a}{\left(\frac{\lambda_0(t)}{1 + \exp(-\boldsymbol{\beta}'\mathbf{z}_j)} \right)^a} \\ &= \left(\frac{1 + \exp(-\boldsymbol{\beta}'\mathbf{z}_j)}{1 + \exp(-\boldsymbol{\beta}'\mathbf{z}_i)} \right)^a \end{aligned} \quad (2.2.3)$$

que no depende del tiempo t y por tanto, el modelo (2.2.1) es un modelo de riesgos proporcionales.

Por otro lado, si se considera $a = 1$ en el modelo (2.2.1) se tiene

$$\lambda(t|\boldsymbol{\beta}, a = 1, \mathbf{z}) = \lambda_0(t) \frac{1}{1 + \exp(-\boldsymbol{\beta}'\mathbf{z})}$$

o lo que es equivalente

$$\lambda(t|\boldsymbol{\beta}, \mathbf{z}) = \lambda_0(t) \frac{\exp(\boldsymbol{\beta}'\mathbf{z})}{\exp(\boldsymbol{\beta}'\mathbf{z}) + 1}. \quad (2.2.4)$$

Esto es, el modelo de riesgos proporcionales con función nexo la función logit. Obsérvese que en MacKenzie (1996, 1997) se propone y analiza el modelo de riesgos no proporcionales usando como función nexo la función logit, por tanto el modelo (2.2.4) es la contraparte a esta propuesta usando un modelo de riesgos proporcionales pero con función de riesgo no constante.

La función de riesgo base acumulativa para el modelo (2.2.1) viene dada por

$$\Lambda_0(t) = \int_0^t \lambda_0(u) du$$

y entonces la función de riesgo acumulativa es

$$\begin{aligned} \Lambda(t|\boldsymbol{\beta}, a, \mathbf{z}) &= \int_0^t \lambda(u|\boldsymbol{\beta}, a, \mathbf{z}) du \\ &= \int_0^t \lambda_0(u) K(\boldsymbol{\beta}, a, \mathbf{z}) du \\ &= K(\boldsymbol{\beta}, a, \mathbf{z}) \int_0^t \lambda_0(u) du \\ &= K(\boldsymbol{\beta}, a, \mathbf{z}) \Lambda_0(t). \end{aligned}$$

Por otra parte, usando la ecuación (1.1.5) la función de supervivencia correspondiente es

$$S(t|\boldsymbol{\beta}, a, \mathbf{z}) = \exp(-K(\boldsymbol{\beta}, a, \mathbf{z}) \Lambda_0(t)) \quad (2.2.5)$$

y por la ecuación (1.1.3), el modelo (2.2.1) es caracterizado por la función de densidad

$$f(t|\boldsymbol{\beta}, a, \mathbf{z}) = \exp\{-K(\boldsymbol{\beta}, a, \mathbf{z}) \Lambda_0(t)\} \lambda_0(t) K(\boldsymbol{\beta}, a, \mathbf{z}).$$

Hasta ahora, no se ha hecho ninguna suposición sobre la función de riesgo base $\lambda_0(\cdot)$. Se puede considerar una función paramétrica y el modelo sería completamente paramétrico o se puede considerar una función arbitraria y el modelo sería semiparamétrico, más adelante se abordaran ambos casos. Ahora, se estudia la interpretación que tienen los parámetros β en el modelo aplicado a datos de supervivencia reales.

2.3. Interpretación del Modelo

Es importante la interpretación de los valores del vector de parámetros β y el efecto que causan éstos en la supervivencia del paciente para valores dados del vector de covariables \mathbf{z} .

Con base en la función de riesgo relativo, se tiene que:

- Si $\rho(\beta, a, \mathbf{z}_i, \mathbf{z}_j) = 1$ los individuos con covariables \mathbf{z}_i y \mathbf{z}_j tienen el mismo riesgo.
- Si $\rho(\beta, a, \mathbf{z}_i, \mathbf{z}_j) < 1$ el individuo con covariable \mathbf{z}_i tiene menor riesgo que el individuo con covariable \mathbf{z}_j .
- Si $\rho(\beta, a, \mathbf{z}_i, \mathbf{z}_j) > 1$ el individuo con covariable \mathbf{z}_i tiene mayor riesgo que el individuo con covariable \mathbf{z}_j .

Por ejemplo, sea el caso en que se tienen datos de supervivencia con una sola covariable z , esta covariable z es la variable indicadora si el paciente recibe un tratamiento A o un tratamiento B. Si $z = 0$, entonces el paciente recibe el tratamiento A y si $z = 1$ el tratamiento B. Se ajusta el modelo de riesgos proporcionales (2.2.1) y se obtienen los valores $\hat{\beta}$ y \hat{a} para estos datos. Si se considera el riesgo relativo ρ dado por (2.2.3), se tiene

$$\rho(\hat{\beta}, \hat{a}, z = 1, z = 0) = \left(\frac{2}{1 + \exp(-\hat{\beta})} \right)^{\hat{a}}.$$

Observe que si $\hat{\beta} = 0$, el riesgo relativo ρ es igual a 1, lo cual significa que ambos tratamientos son igual de efectivos. Si $\hat{\beta} > 0$ el individuo que recibe el tratamiento B tiene menor riesgo que el individuo que recibe el tratamiento A (el tratamiento B es más efectivo que el tratamiento A), y viceversa si $\hat{\beta} < 0$ el individuo que recibe el tratamiento B tiene mayor riesgo que el individuo que recibe el tratamiento A (el tratamiento A es más efectivo que el tratamiento B).

En general, es de especial interés estimar el riesgo relativo cuando aumenta la covariable z_m en una unidad y el resto de covariables se mantienen constantes, para ello se toma $\mathbf{z}_j = (z_{j1}, \dots, z_{jm}, \dots, z_{jd})$ y $\mathbf{z}_i = (z_{j1}, \dots, z_{jm} + 1, \dots, z_{jd})$ entonces se quiere saber bajo que condiciones el riesgo relativo $\rho(\boldsymbol{\beta}, a, \mathbf{z}_i, \mathbf{z}_j)$ es mayor, igual o menor que uno.

Sea el caso para el cual $\rho(\boldsymbol{\beta}, a, \mathbf{z}_i, \mathbf{z}_j) < 1$, entonces

$$\begin{aligned} \left(\frac{1 + \exp(-\boldsymbol{\beta}'\mathbf{z}_j)}{1 + \exp(-\boldsymbol{\beta}'\mathbf{z}_i)} \right)^a < 1 &\Leftrightarrow \frac{1 + \exp(-\boldsymbol{\beta}'\mathbf{z}_j)}{1 + \exp(-\boldsymbol{\beta}'\mathbf{z}_i)} < 1 \\ &\Leftrightarrow 1 + \exp(-\boldsymbol{\beta}'\mathbf{z}_j) < 1 + \exp(-\boldsymbol{\beta}'\mathbf{z}_i) \\ &\Leftrightarrow \exp(-\boldsymbol{\beta}'\mathbf{z}_j) < \exp(-\boldsymbol{\beta}'\mathbf{z}_i) \\ &\Leftrightarrow \exp[\boldsymbol{\beta}'(\mathbf{z}_i - \mathbf{z}_j)] < 1 \\ &\Leftrightarrow \exp(\beta_m) < 1 \\ &\Leftrightarrow \beta_m < 0. \end{aligned}$$

Análogamente,

$$\rho(\boldsymbol{\beta}, a, \mathbf{z}_i, \mathbf{z}_j) > 1 \text{ si y sólo si } \beta_m > 0$$

y

$$\rho(\boldsymbol{\beta}, a, \mathbf{z}_i, \mathbf{z}_j) = 1 \text{ si y sólo si } \beta_m = 0.$$

Por lo anterior, se puede decir que si un individuo aumenta en una unidad el valor de la covariable z_m y mantiene fijos el resto tendrá menor (o mayor) riesgo si el respectivo parámetro es negativo (o positivo), o el mismo riesgo si el parámetro es cero.

Ahora se sabe como se interpretan los parámetros del modelo (2.2.1), sin embargo, no se ha estudiado como estimar dichos parámetros o lo que es lo mismo, ajustar el modelo (2.2.1) a un conjunto de datos de supervivencia. En las secciones siguientes se estudia la estimación de los parámetros cuando la función de riesgo es una función que depende de un número finito de parámetros (caso paramétrico) y cuando es una función arbitraria (caso semiparamétrico). No obstante, en primer lugar, se describe el método para generar datos de supervivencia del modelo, ya que será utilizado en la última sección del capítulo.

2.4. Simulación de Tiempos de Supervivencia

Como se mencionó en el Capítulo 1, algunas veces es útil generar tiempos de supervivencia provenientes de un modelo particular, en este caso el modelo generalizado de riesgos proporcionales (2.2.1).

Para generar el tiempo de supervivencia t se puede adaptar la metodología propuesta por Bender y otros (2005), descrita en la Sección 1.2.3. En este caso, se sustituye la expresión de la función de supervivencia del modelo generalizado de riesgos proporcionales (2.2.5) en la expresión (1.2.24) obteniendo que

$$u = \exp(-K(\boldsymbol{\beta}, a, \mathbf{z})\Lambda_0(t)),$$

donde $u \sim \mathcal{U}(0, 1)$. Si la función $\Lambda_0(\cdot)$ es invertible se tiene que el tiempo t está dado por

$$t = \Lambda_0^{-1}\left(-\frac{\log u}{K(\boldsymbol{\beta}, a, \mathbf{z})}\right).$$

Como caso particular, si la función de riesgo base $\lambda_0(\cdot)$ es una función Weibull con parámetros γ y ν se tiene que el tiempo de supervivencia t se obtiene usando la siguiente expresión

$$t = \left(-\frac{1}{\gamma} \frac{\log u}{K(\boldsymbol{\beta}, a, \mathbf{z})}\right)^{\frac{1}{\nu}}, \quad (2.4.6)$$

con $u \sim \mathcal{U}(0, 1)$.

Más adelante, se generarán datos de supervivencia con la finalidad de observar y/o verificar las propiedades de los estimadores propuestos para el modelo generalizado de riesgos proporcionales.

2.5. Modelo Paramétrico

2.5.1. Definición del Modelo

En análisis de supervivencia, algunas veces se usan modelos paramétricos, particularmente para modelar la función de riesgo, algunos modelos ampliamente usados son: el exponencial, el Weibull, el gamma, el log-normal, el log-logistic, el normal, el potencia exponencial, el Gompertz, el Gaussiano inverso, el Pareto y el gamma generalizado. En la Tabla 2.2 de Klein y Moeschberger (2003) se resume la información de estos modelos, ya que son muy utilizados por investigadores médicos.

En esta sección, se asume que la función de riesgo base es una función paramétrica conocida. En particular, como ejemplo, supóngase una distribución Weibull con parámetros γ y μ , que llevaría a un modelo Weibull logit paramétrico generalizado de riesgos proporcionales dado por la función de riesgo

$$\lambda(t|\mu, \gamma, \boldsymbol{\beta}, a, \mathbf{z}) = \mu\gamma t^{\gamma-1} K(\boldsymbol{\beta}, a, \mathbf{z}) \quad (2.5.7)$$

con $\mu, \gamma, a > 0$. Además, la función de riesgo acumulativa viene dada por

$$\Lambda(t|\mu, \gamma, \boldsymbol{\beta}, a, \mathbf{z}) = \mu t^\gamma K(\boldsymbol{\beta}, a, \mathbf{z}),$$

y el modelo (2.5.7) queda caracterizado por la función de densidad

$$\begin{aligned} f(t|\mu, \gamma, \boldsymbol{\beta}, a, \mathbf{z}) &= \exp\{-K(\boldsymbol{\beta}, a, \mathbf{z})\mu t^\gamma\} \mu\gamma t^{\gamma-1} K(\boldsymbol{\beta}, a, \mathbf{z}) \\ &= \mu K(\boldsymbol{\beta}, a, \mathbf{z}) \gamma t^{\gamma-1} \exp\{-\mu K(\boldsymbol{\beta}, a, \mathbf{z}) t^\gamma\}. \end{aligned}$$

Es decir, una función de densidad Weibull con parámetros γ y $\mu K(\beta, a, \mathbf{z})$. Además, por la ecuación (2.2.5) la función de supervivencia para este modelo es

$$S(t|\mu, \gamma, \beta, a, \mathbf{z}) = \exp \{-K(\beta, a, \mathbf{z})\mu t^\gamma\}. \quad (2.5.8)$$

Tomando $\gamma = 1$ en el modelo (2.5.7), se tiene el modelo Exponencial logit generalizado de riesgos proporcionales.

Por otro lado, nótese que, además de los parámetros de la distribución Weibull, para el modelo propuesto se tiene un parámetro de regresión más que en el modelo de regresión de Cox, el parámetro a . Este parámetro dará flexibilidad en el ajuste del modelo a un conjunto de datos. En la Figura 2.1, se muestra la función de supervivencia y la función riesgo para diferentes valores de a manteniendo fijos los valores de los parámetros $\mu = 1.0$ y $\gamma = 0.5, 1.5, 3.0$, respectivamente, además se considera $\beta = 1$ y $z = 1$.

2.5.2. Estimador de los Parámetros y sus Propiedades Asintóticas

A continuación, se describe el método de máxima verosimilitud para estimar los parámetros del modelo Weibull logit de riesgos proporcionales (2.5.7), es decir, el modelo generalizado de riesgos proporcionales (2.2.1) cuando la función de riesgo base es paramétrica (Weibull), considerando datos con censura por la derecha.

Se consideran n tripletas iid (T_i, \mathbf{Z}_i, C_i) con posible censura por la derecha, donde T_i y C_i son el tiempo de supervivencia y censura respectivamente y \mathbf{Z}_i es el vector de covariables. Se asume que ambos tiempos T_i y C_i están en el intervalo $[0, \tau]$ y que el tiempo T_i es generado bajo el modelo (2.5.7) dado \mathbf{Z}_i y condicionalmente independiente de C_i y que éste no depende \mathbf{Z}_i . Entonces los datos observados son de la forma $(X_i, \delta_i, \mathbf{Z}_i)$ con $i = 1, \dots, n$ donde $X_i = \min(T_i, C_i)$ y $\delta_i = \mathbb{1}_{\{X_i = C_i\}}$.

De acuerdo con el mecanismo no informativo de censura, la función de verosi-

Modelo Generalizado para Datos con Censura por la Derecha

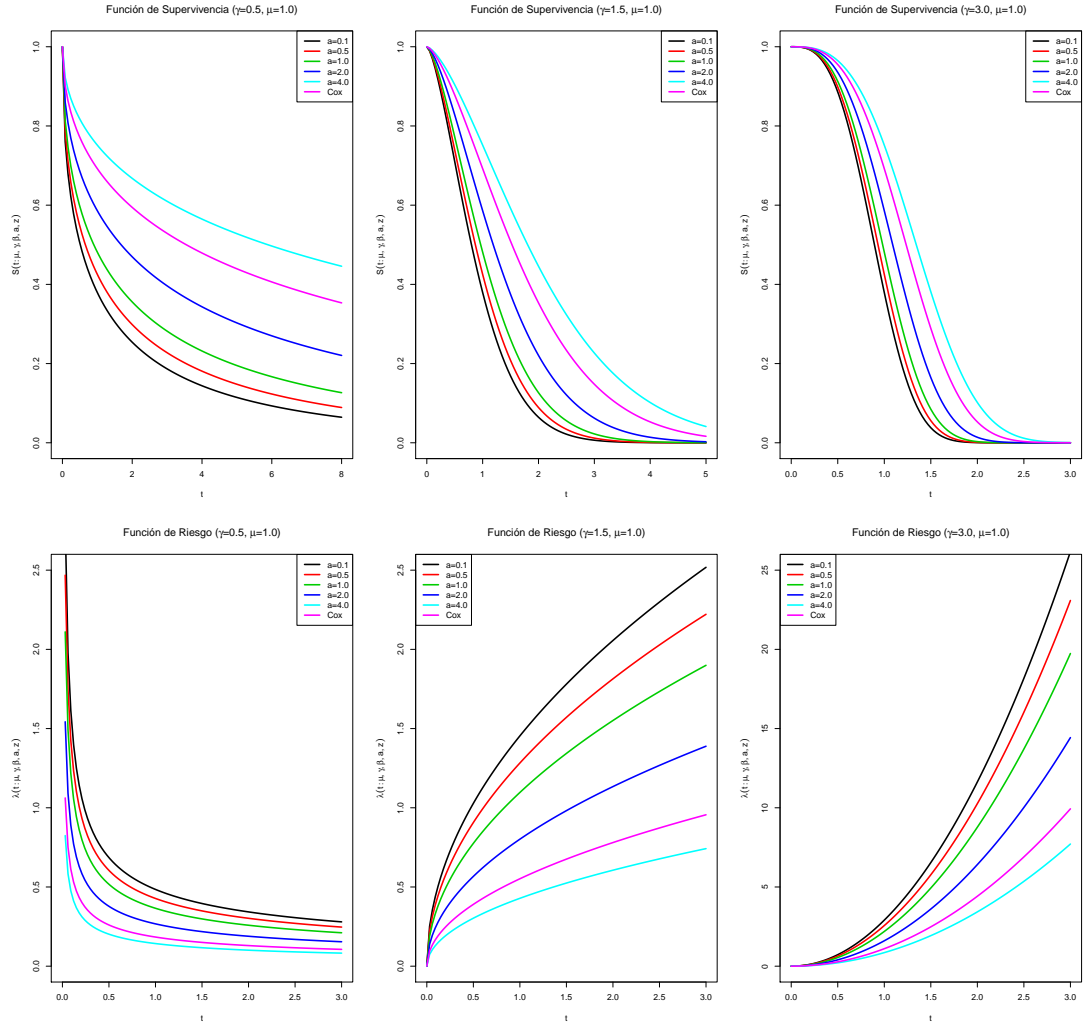


Figura 2.1: Gráficas de la función de supervivencia y la función de riesgo del modelo Weibull logit generalizado para diferentes valores de a .

militud de la ecuación (1.1.9) se puede escribir como

$$\begin{aligned}
 \mathcal{L}_n(\mu, \gamma, \boldsymbol{\beta}, a) &= \prod_{i=1}^n [f(x_i|\mu, \gamma, \boldsymbol{\beta}, a, \mathbf{z}_i)]^{\delta_i} [S(x_i|\mu, \gamma, \boldsymbol{\beta}, a, \mathbf{z}_i)]^{1-\delta_i} \\
 &= \prod_{i=1}^n [\lambda(x_i|\mu, \gamma, \boldsymbol{\beta}, a, \mathbf{z}_i)S(x_i|\mu, \gamma, \boldsymbol{\beta}, a, \mathbf{z}_i)]^{\delta_i} \\
 &\quad [S(x_i|\mu, \gamma, \boldsymbol{\beta}, a, \mathbf{z}_i)]^{1-\delta_i} \\
 &= \prod_{i=1}^n [\lambda(x_i|\mu, \gamma, \boldsymbol{\beta}, a, \mathbf{z}_i)]^{\delta_i} S(x_i|\mu, \gamma, \boldsymbol{\beta}, a, \mathbf{z}_i) \\
 &= \prod_{i=1}^n [\mu\gamma x_i^{\gamma-1} K(\boldsymbol{\beta}, a, \mathbf{z}_i)]^{\delta_i} \exp\{-K(\boldsymbol{\beta}, a, \mathbf{z}_i)\mu x_i^\gamma\}
 \end{aligned}$$

entonces la función de log-verosimilitud para el modelo (2.5.7) está dada por

$$\begin{aligned}
 \ell_n(\mu, \gamma, \boldsymbol{\beta}, a) &= \log \mathcal{L}_n(\mu, \gamma, \boldsymbol{\beta}, a) \\
 &= \sum_{i=1}^n [\delta_i \log(\mu\gamma x_i^{\gamma-1} K(\boldsymbol{\beta}, a, \mathbf{z}_i)) - K(\boldsymbol{\beta}, a, \mathbf{z}_i)\mu x_i^\gamma]. \quad (2.5.9)
 \end{aligned}$$

Así, los parámetros estimados serán

$$(\hat{\mu}_n, \hat{\gamma}_n, \hat{\boldsymbol{\beta}}_n, \hat{a}_n) = \overline{\text{arg}} \max_{(\mu, \gamma, \boldsymbol{\beta}, a)} \ell_n(\mu, \gamma, \boldsymbol{\beta}, a) \quad (2.5.10)$$

y

$$\hat{\Lambda}_0(t|\hat{\mu}_n, \hat{\gamma}_n) = \hat{\mu}_n \hat{\gamma}_n t^{\hat{\gamma}_n-1}.$$

Por otro lado, como los estimadores (2.5.10) son estimadores de máxima verosimilitud se pueden denotar por $\hat{\boldsymbol{\theta}}_n = (\hat{\mu}_n, \hat{\gamma}_n, \hat{\boldsymbol{\beta}}_n, \hat{a}_n)$ y se tienen todas las propiedades de éstos como consistencia y normalidad asintótica bajo las condiciones habituales de la función log-verosimilitud, es decir

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0$$

cuando $n \rightarrow \infty$ donde $\boldsymbol{\theta}_0 = (\mu_0, \gamma_0, \boldsymbol{\beta}_0, a_0)$ es el verdadero valor de parámetro. Además,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(\mathbf{0}, \mathbf{I}^{-1})$$

donde \mathbf{I} es la matriz de información, es decir es la matriz hessiana de la función (2.5.9).

En este caso, para encontrar el máximo de la función de log-verosimilitud de la función (2.5.9) se usan métodos numéricos para optimización con restricciones, ya que como se sabe que $\mu, \gamma, a > 0$, estos métodos numéricos usan la función gradiente de (2.5.9).

En el Apéndice I, se obtienen las primeras y segundas derivadas de la función de log-verosimilitud (2.5.9), para obtener el vector gradiente y la matriz hessiana de la función (2.5.9), respectivamente, necesarios para la optimización de la función de log-verosimilitud y en la estimación del error estándar de los estimadores.

2.5.3. Aplicación a Datos Reales

Estudio de Cáncer de Laringe

Se ilustra el modelo paramétrico propuesto en este capítulo con los datos estudiados por Kardaun (1983), los cuales son tiempos de muerte causada por cáncer de laringe. La base de datos cuenta con 90 registros de hombres diagnosticados con cáncer de laringe entre 1970 y 1978. El tiempo registrado es el intervalo (en años) entre el primer tratamiento y la muerte o el final del estudio. También se dispone de la edad de diagnóstico (Z_1) y la etapa de la enfermedad en que se encuentra cada paciente, ésta última es una covariable tipo factor con 4 niveles (es decir, $z_2 = 1$ si el paciente se encuentra en la etapa II de la enfermedad o $z_2 = 0$ en caso contrario, $z_3 = 1$ si el paciente se encuentra en el etapa III de la enfermedad o $z_3 = 0$ en caso contrario y $z_4 = 1$ si el paciente se encuentra en el etapa IV de la enfermedad o $z_4 = 0$ en caso contrario). Estos datos también los utilizan Klein y Moeschberger (2003) para ilustrar diversas técnicas de análisis de supervivencia. Los datos están disponibles con el nombre `larynx` en el paquete `MKsurv` del software libre R, y se presentan en el Apéndice III de este trabajo.

Se ajustan los modelos de riesgos proporcionales con función de riesgo base una función Weibull y con función nexa la función exponencial (modelo de Cox

paramétrico), la función logit y la función logit generalizada, que vienen dados por la expresiones:

$$\text{WPH: } \lambda(t|\mu, \gamma, \boldsymbol{\beta}, \mathbf{z}) = \mu\gamma t^{\gamma-1} \exp(\boldsymbol{\beta}'\mathbf{z}),$$

$$\text{WLPH: } \lambda(t|\mu, \gamma, \boldsymbol{\beta}, \mathbf{z}) = \mu\gamma t^{\gamma-1} K(\boldsymbol{\beta}, 1, \mathbf{z}),$$

$$\text{WGLPH: } \lambda(t|\mu, \gamma, \boldsymbol{\beta}, a, \mathbf{z}) = \mu\gamma t^{\gamma-1} K(\boldsymbol{\beta}, a, \mathbf{z}),$$

respectivamente, con $K(\boldsymbol{\beta}, a, \mathbf{z})$ como en (2.2.2). Estos modelos se denotan por WPH, WLPH y WGLPH por sus siglas en inglés.

Cabe mencionar que al ajustar el modelo WGLPH se observa que el algoritmo de optimización usado para maximizar la función de log-verosimilitud es muy sensible a valores iniciales para el parámetro a , debido a esto se estanca en óptimos locales. Por lo anterior, se ha optado por el método de maximizar la verosimilitud por contornos, es decir, se busca el máximo de $\ell_n(\mu, \gamma, \boldsymbol{\beta}, a)$, en dos etapas de modo que, en la primera etapa se maximiza la función ℓ_n manteniendo fijo el valor del parámetro a , con a en un conjunto de valores preestablecidos $\{a_1, a_2, \dots, a_r\}$ y, en la segunda etapa, se selecciona la función

$$\ell_n(\mu, \gamma, \boldsymbol{\beta}, a_1), \ell_n(\mu, \gamma, \boldsymbol{\beta}, a_2), \dots, \ell_n(\mu, \gamma, \boldsymbol{\beta}, a_r),$$

con valor máximo. Finalmente, los estimadores de los parámetros de esta función de log-verosimilitud son los estimadores de los parámetros desconocidos del modelo. El modelo ajustado usando maximización por contornos se denotará por WGLPHaF. Para estos datos se ha considerado $a \in \{5, 6, 7, 8, 9\}$, ya que se ha supuesto que el valor de \hat{a} maximizando por contornos sería cercano al valor de \hat{a} del modelo WGLPH, obteniendo así, el máximo en $\hat{a} = 8$.

Se hace una comparación de los modelos obteniendo la estimación de los parámetros y su respectivo error estándar (EE), además de usar el criterio de información de Akaike y Bayesiano denotados por AIC y BIC por sus siglas en inglés.

La estimación de los parámetros y el error estándar para cada modelo se muestra en la Tabla 2.1, donde se puede ver que los errores estándar para los efectos de las

Modelo Generalizado para Datos con Censura por la Derecha

Covariable	WPH		WLPH		WGLPHaF ($\hat{a} = 8$)		WGLPH	
	Est	EE	Est	EE	Est	EE	Est	EE
μ	0.01855	0.01893	0.32468	0.11235	3.16696	4.94242	2.56534	6.59244
γ	1.13020	0.13843	1.09364	0.13413	1.13038	0.13832	1.13034	0.13847
edad	0.01973	0.01420	-0.01638	0.00810	0.00740	0.00751	0.00785	0.00755
etapa II	0.16633	0.46111	-0.04115	0.60885	0.04320	0.15510	0.04406	0.16261
etapa III	0.66255	0.35545	0.75535	0.56191	0.23229	0.16294	0.24550	0.21569
etapa IV	1.74530	0.41471	11.26756	96.80031	0.73511	0.41099	0.79024	0.75908
a	—	—	—	—	—	—	7.74186	3.94820

Tabla 2.1: Estimadores y errores estándar para los cuatro modelos paramétricos

	WPH	WLPH	WGLPHaF	WGLPH
AIC	294.8468	300.4143	294.7641	296.7652
BIC	309.8456	315.4131	309.7630	314.2639

Tabla 2.2: Valores de AIC y BIC para los cuatro modelos paramétricos

covariables son menores para los estimadores del modelo WGLPHaF. En la Tabla 2.2 se comparan los modelos en base al AIC y al BIC y como se puede observar el modelo WGLPHaF tiene menor valor en ambos criterios, por tanto, en base a estas tablas se puede decir que el modelo WGLPHaF es una buena alternativa al modelo clásico paramétrico de Cox.

En la Figura 2.2, se muestra la estimación de la función de supervivencia, $\hat{S}(t|\hat{\mu}, \hat{\gamma}, \hat{\beta}, \hat{a}, \mathbf{z})$ dada por (2.5.8) para los cuatro modelos ajustados y se comparan con el estimador de Kaplan–Meier. Se observa que las funciones de supervivencia son similares excepto para el modelo con función nexo la función logit, es decir $a = 1$. Esto muestra la flexibilidad que da el parámetro a al modelo propuesto.

Por último, es interesante la comparación de los modelos WPH y WGLPHaF en base al riesgo relativo y a la interpretación de sus parámetros, como se hace en el Ejemplo 8.2. de Klein y Moeschberger (2003). Si se compara un individuo de 50 años de edad con uno de 40 años de edad, con ambos individuos en la misma etapa de la enfermedad. En principio, dado que el estimador del parámetro $\hat{\beta}_1 > 0$ para

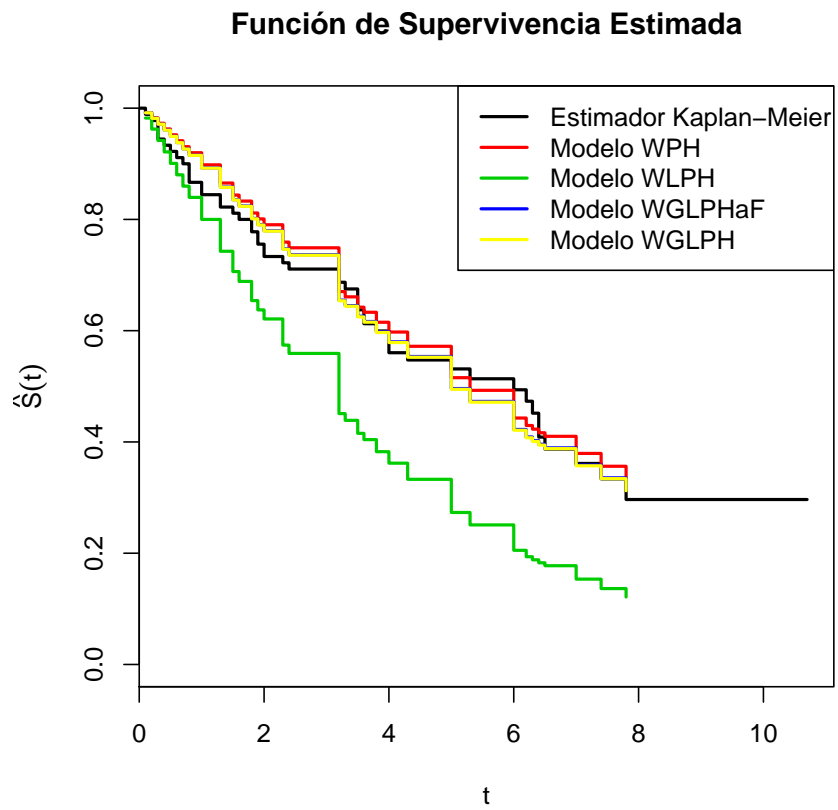


Figura 2.2: Gráfica de la función de supervivencia estimada para los 4 modelos y el estimador de Kaplan-Meier

ambos modelos, se sabe que el individuo de 50 años tiene mayor riesgo de morir para ambos modelos. Para el modelo WPH, se tiene que el riesgo relativo es

$$\exp(10\hat{\beta}_1) \approx 1.22,$$

es decir, el paciente de 50 años de edad comparado con uno de 40 años tiene un riesgo de morir 1.22 veces mayor, independientemente de la etapa de la enfermedad en la que se encuentren siempre y cuando se encuentren en la misma etapa de la enfermedad. En cambio, para el modelo WGLPHaF el valor del riesgo relativo cambia dependiendo de la etapa en que se encuentre el paciente. Supóngase que se comparan individuos que en encuentran en la etapa II entonces el riesgo relativo es

$$\left(\frac{1 + \exp(-40\hat{\beta}_1 - \hat{\beta}_2)}{1 + \exp(-50\hat{\beta}_1 - \hat{\beta}_2)} \right)^{\hat{a}} \approx 1.27.$$

Análogamente, si ambos individuos se encuentran en las etapas III y IV el riesgo es 1.24 y 1.16, respectivamente. En el caso de que ambos individuos estén en la etapa I de la enfermedad el riesgo relativo es

$$\left(\frac{1 + \exp(-40\hat{\beta}_1)}{1 + \exp(-50\hat{\beta}_1)} \right)^{\hat{a}} \approx 1.28.$$

Esto es, el paciente de 50 años de edad tiene un riesgo mayor de morir que uno de 40 años, dado que ambos individuos están en la misma etapa de la enfermedad, sin embargo, este valor es diferente para cada etapa, el riesgo relativo decrece a medida que la enfermedad se encuentra en una etapa superior. Es decir, que cuando la enfermedad se encuentra en una etapa avanzada la edad del paciente va perdiendo importancia. Esta información es valiosa y no se puede obtener con el modelo clásico de Cox.

Posteriormente, se retomará este ejemplo para comparar la familia de modelos semiparamétrica propuesta y los diferentes estimadores de la función de supervivencia que se propondrán para ellos.

2.6. Modelo Semiparamétrico

Supóngase que n es el número de individuos en el estudio, asociado a cada individuo se tiene el vector aleatorio $\mathbf{Y}_i = (X_i, \delta_i, \mathbf{Z}_i)$ donde $X_i = \min(T_i, C_i)$, y $\delta_i = \mathbb{1}_{\{X_i=C_i\}}$ con $i = 1, \dots, n$, que se suponen vectores aleatorios independientes e idénticamente distribuidos de datos de supervivencia con censura por la derecha.

Se considera el modelo (2.2.1), pero en este caso se asume que la función de riesgo base $\lambda_0(\cdot)$ es una función arbitraria desconocida, es decir, ahora se tiene el modelo semiparamétrico y se debe estimar el vector de parámetros $\boldsymbol{\beta}$, el parámetro a y la función de riesgo acumulativa $\Lambda_0(\cdot)$. Con la finalidad de estimar los parámetros $\boldsymbol{\beta}$ y a se maximiza la función de verosimilitud parcial y posteriormente, se considera la función de verosimilitud “profile” con el objetivo de demostrar la normalidad asintótica de dichos parámetros. Finalmente, en la Sección 2.6.5, se usará la función de verosimilitud total para obtener un estimador de la función $\Lambda_0(\cdot)$.

2.6.1. Estimación de los Parámetros

Para obtener una estimación de $\boldsymbol{\beta}$ y a en base a la función de verosimilitud parcial, se puede seguir la misma metodología que Cox (1972) y que se ha descrito detalladamente en la Sección 1.2 concretamente en 1.2.1, con la diferencia que se sustituye la forma del modelo (2.2.1) en la ecuación (1.2.18). Así, se obtiene que la verosimilitud parcial para un dato no censurado j es

$$\frac{\lambda_0(t_j)K(\boldsymbol{\beta}, a, \mathbf{z}_j)}{\sum_{\ell \in R(t_j)} \lambda_0(t_j)K(\boldsymbol{\beta}, a, \mathbf{z}_\ell)} = \frac{K(\boldsymbol{\beta}, a, \mathbf{z}_j)}{\sum_{\ell \in R(t_j)} K(\boldsymbol{\beta}, a, \mathbf{z}_\ell)}$$

y la expresión correspondiente a la función de verosimilitud parcial para los n datos viene dada por

$$\mathcal{L}_n(\boldsymbol{\beta}, a) = \prod_{i=1}^n \left(\frac{K(\boldsymbol{\beta}, a, \mathbf{z}_i)}{\sum_{\ell \in R(x_i)} K(\boldsymbol{\beta}, a, \mathbf{z}_\ell)} \right)^{\delta_i}.$$

Por tanto, la función de log-verosimilitud parcial es

$$\begin{aligned} \ell_n(\boldsymbol{\beta}, a) &= \log \ell_n(\boldsymbol{\beta}, a) \\ &= \sum_{i=1}^n \delta_i \left(\log K(\boldsymbol{\beta}, a, \mathbf{z}_i) - \log \sum_{\ell \in R(X_i)} K(\boldsymbol{\beta}, a, \mathbf{z}_\ell) \right). \end{aligned} \quad (2.6.11)$$

Así, los parámetros estimados son

$$(\hat{\boldsymbol{\beta}}_n, \hat{a}_n) = \arg \max_{(\boldsymbol{\beta}, a > 0)} \ell_n(\boldsymbol{\beta}, a). \quad (2.6.12)$$

Para resolver el problema anterior de optimización con restricciones es necesario utilizar algún método numérico, por ejemplo métodos del tipo quasi-Newton. En el Apéndice I, se obtienen las primeras y segundas derivadas parciales de la función (2.6.11), ya que se necesitarán en el procedimiento de optimización numérica y para estimar el error estándar de los parámetros estimados, respectivamente.

A continuación, se demuestra la normalidad asintótica del vector de parámetros estimados $\hat{\boldsymbol{\beta}}$, además de proponer un estimador $\hat{\Lambda}_0(\cdot)$ para la función acumulativa de riesgo. No sin antes, aplicar el método anterior a un conjunto de datos de supervivencia reales.

2.6.2. Aplicación a Datos Reales

Estudio de Cáncer de Laringe

Como en la Sección 2.5.3, para ilustrar el uso del modelo semiparamétrico generalizado de riesgos proporcionales se usan los datos `larynx`. Se ajustan los modelos semiparamétricos de riesgos proporcionales con función nexa la función exponencial, la función logit y la función logit generalizada denotados por PH, LPH, y GLPH por sus siglas en inglés, respectivamente. También se considera el caso donde el parámetro a se estima siguiendo la metodología de maximización por contornos, descrito en la Sección 2.5.3, con la finalidad de evitar que el algoritmo de optimización se estanque en máximos locales cercanos al valor inicial del parámetro a , a este

Modelo Generalizado para Datos con Censura por la Derecha

Covariable	PH		LPH		GLPHaF ($\hat{a} = 8$)		GLPH	
	Est	EE	Est	EE	Est	EE	Est	EE
edad	0.01890	0.01425	-0.01568	0.00828	0.00570	0.00565	0.00574	0.02170
etapa II	0.13856	0.46231	-0.07464	0.61227	0.03047	0.12829	0.03060	0.14502
etapa III	0.63835	0.35608	0.75403	0.57134	0.18337	0.12140	0.18453	0.67040
etapa IV	1.69306	0.42221	7.93424	19.34429	0.55128	0.23583	0.55537	2.35490
a	—	—	—	—	—	—	8.95955	22.95138

Tabla 2.3: Estimadores y errores estándar de los cuatro modelos semiparamétricos

	PH	LPH	GLPHaF	GLPH
AIC	384.3589	389.2912	384.3173	386.3173
BIC	394.3581	399.2905	394.3165	398.8163

Tabla 2.4: Valores de AIC y BIC para los cuatro modelos semiparamétricos

modelo lo denotamos por GLPHaF. En este caso, se ha tomado $a \in \{1, 2, \dots, 9, 10\}$, obteniendo un valor máximo para la función de log-verosimilitud en $\hat{a} = 8$.

En la Tabla 2.3 se comparan los modelos en base al error estándar (EE) y en la Tabla 2.4 se comparan en base a los criterios de información de Aikake y Bayesiano (AIC y BIC). Se puede notar que para el modelo GLPHaF la estimación de valores de los parámetros tienen menor EE estimado. Además, de acuerdo a los criterios AIC y BIC, también se obtienen los valores más pequeños para el modelo GLPHaF. Por tanto, el modelo semiparamétrico GLPHaF vuelve a destacar como el mejor.

Finalmente, como en la Sección 2.5.3, se comparan los modelos PH y GLPHaF en base al riesgo relativo obtenido con cada modelo. Como en la sección anterior, se consideran un individuo de 50 años de edad y uno de 40 años de edad con ambos individuos en la misma etapa de la enfermedad. Para el modelo PH el riesgo relativo es

$$\exp(10\hat{\beta}_1) \approx 1.21,$$

para individuos en la misma etapa de la enfermedad pero cualquiera de las cuatro etapas, es decir, un individuo de 50 años tiene un riesgo de morir 1.21 veces mayor

que uno de 40 años que se encuentra en la misma etapa de enfermedad que él. En cambio, para el modelo GLPHaF se tiene un riesgo relativo diferente para cada etapa, esto es, para estos individuos en la etapa I se tiene un riesgo relativo igual a

$$\left(\frac{1 + \exp(-40\hat{\beta}_1)}{1 + \exp(-50\hat{\beta}_1)} \right)^{\hat{a}} \approx 1.22,$$

en la etapa II se tiene un riesgo de

$$\left(\frac{1 + \exp(-40\hat{\beta}_1 - \hat{\beta}_2)}{1 + \exp(-50\hat{\beta}_1 - \hat{\beta}_2)} \right)^{\hat{a}} \approx 1.22,$$

análogamente, en la etapa III y IV el riesgo es aproximadamente 1.20 y 1.15, respectivamente. Entonces, el individuo de 50 años de edad tiene mayor riesgo de morir que el de 40, pero si se encuentran en la etapa I y II el riesgo es 1.22 veces mayor, si se encuentran en la etapa III es 1.2 veces mayor y si están en la etapa IV es 1.15. Es decir, que el modelo logit generalizado estima un riesgo relativo decreciente de muerte a medida que la enfermedad avanza en gravedad. Esta conclusión importante no se puede obtener con el modelo de Cox.

Por otra parte, notése que para el modelo PH el riesgo relativo para un individuo comparado con otro 10 años menor, siempre es el mismo. En cambio para el modelo GLPHaF no es así, por ejemplo, supóngase que comparamos un individuo de 50 años con uno de 40 años, ambos en la etapa II de la enfermedad, como se sabe el primer individuo tiene un riesgo de morir 1.22 veces mayor que el de 40 años, sin embargo si comparamos un individuo de 60 años con uno de 50 años, ambos en la etapa II de la enfermedad, el individuo de 60 años tiene un riesgo

$$\left(\frac{1 + \exp(-50\hat{\beta}_1 - \hat{\beta}_2)}{1 + \exp(-60\hat{\beta}_1 - \hat{\beta}_2)} \right)^{\hat{a}} \approx 1.21$$

veces mayor de morir que el de 50. Análogamente, si comparamos uno de 80 años con uno de 70 años, el primero tiene un riesgo 1.19 veces mayor de morir que el segundo. Esto es, con el modelo GLPHaF se obtienen riesgos relativos específicos

para cada rango de edad. En concreto, se puede concluir que para la misma diferencia de edad el riesgo relativo de muerte disminuye cuando la edad del paciente crece.

Estudio de Cáncer de Pulmón

Por otro lado, también se analizan los datos estudiados por MacKenzie (1996) que nos han sido proporcionados por el autor. Los datos son tiempos de supervivencia de un estudio clínico sobre la incidencia de cáncer de pulmón, el estudio se llevó a cabo entre Octubre 1 de 1991 y Septiembre 30 de 1992, en el norte de Irlanda, durante este periodo se diagnosticaron 900 casos. Para cada individuo se tiene información clínica adicional. La base de datos cuenta con 855 datos completos (95 %), 20 datos para los cuales se hizo el diagnóstico post-mortem y 25 datos no determinados.

Al igual que MacKenzie (1996), se consideran las covariables: edad de diagnóstico (en años) y sexo (1=masculino y 0=femenino). El tiempo de supervivencia es el tiempo entre el primer diagnóstico y la muerte o censura. Se ajustan los modelos PH, LPH y GLPHaF, ya que como con los datos `larynx` ajustando el modelo GLPHaF en vez de GLPH, se evita el estancamiento en óptimos locales cercanos al valor inicial del parámetro a .

En la Tabla 2.5 se muestran las estimaciones de los parámetros y su correspondiente error estándar (EE), se puede notar que los modelos PH y GLPHaF son similares. No obstante, en la Tabla 2.6 se comparan los modelos en base al criterio de información de Akaike y Bayesiano y se puede concluir que el modelo generalizado de riesgos proporcionales (GLPHaF) tiene mejor ajuste para los datos, ya que los valores de AIC y BIC son menores para dicho modelo.

Por otra parte, se comparan los modelos PH y GLPHaF en base al riesgo relativo obtenido con cada modelo. Se consideran un individuo de 70 años de edad y un

Modelo Generalizado para Datos con Censura por la Derecha

Modelo	Covariable	Est	EE
PH	Edad	0.017	0.004
	Sexo	0.017	0.082
LPH	Edad	0.020	0.010
	Sexo	-0.034	0.385
GLPHaF ($\hat{a}=4.5$)	Edad	0.014	0.010
	Sexo	0.024	0.072

Tabla 2.5: Estimadores y errores estándar de los tres modelos semiparamétricos

	PH	LPH	GLPHaF
AIC	8194.797	8205.509	8194.693
BIC	8204.300	8215.011	8204.195

Tabla 2.6: Valores de AIC y BIC para los tres modelos semiparamétricos

individuo de 60 años de edad, ambos individuos del mismo sexo. Considerando el modelo PH, el riesgo relativo para este caso es

$$\exp(10\hat{\beta}_1) \approx 1.19,$$

esto es, un individuo de 70 años tiene un riesgo relativo de morir 1.19 veces mayor que un individuo de 60 años, ambos individuos del mismo sexo. En cambio, para obtener el riesgo relativo asumiendo el modelo GLPHaF, primero supóngase que los individuos considerados son mujeres, entonces el riesgo relativo es

$$\left(\frac{1 + \exp(-60\hat{\beta}_1)}{1 + \exp(-70\hat{\beta}_1)} \right)^{\hat{a}} \approx 1.20,$$

en cambio, si se consideran individuos hombres, el riesgo relativo es

$$\left(\frac{1 + \exp(-60\hat{\beta}_1 - \hat{\beta}_2)}{1 + \exp(-70\hat{\beta}_1 - \hat{\beta}_2)} \right)^{\hat{a}} \approx 1.19.$$

Esto es, el modelo logit generalizado estima un mayor riesgo relativo en mujeres que en hombres, en cambio el modelo de Cox estima el mismo riesgo relativo para ambos sexos. Por último, como sabemos el riesgo relativo de individuos del mismo sexo

con diferencia de 10 años, bajo el modelo de Cox es siempre el mismo. En cambio, no es así si se considera el modelo generalizado, por ejemplo, si los individuos son mujeres pero comparamos un individuo de 50 años con uno de 40 años, uno de 60 años con uno de 50 años y uno de 70 años con uno de 60 años se obtienen riesgos relativos de 1.24, 1.22 y 1.20, respectivamente. Análogamente, si comparamos el riesgo relativo en hombres se tiene que son 1.24, 1.22 y 1.19, respectivamente. De donde se concluye que el riesgo relativo con la misma diferencia de edad disminuye cuando la edad del paciente aumenta. Sin embargo, la diferencia en el riesgo entre el grupo de mujeres y hombres es mínima y se pierde con el redondeo del cálculo numérico.

2.6.3. Propiedades Asintóticas de los Estimadores de los Parámetros

Para demostrar la normalidad asintótica de los estimadores de los parámetros del modelo semiparamétrico logit generalizado se aplicará el resultado de Murphy y van der Vaart (2000) enunciado como Teorema 1.3.10 en el Capítulo 1 de este trabajo. Es decir, es necesario definir un submodelo menos favorable y verificar que dicho submodelo satisface las condiciones (1.3.31) y (1.3.32). Por otro lado, los estimadores deben cumplir las condiciones (1.3.33) y (1.3.34). Después, es necesario verificar las hipótesis del Teorema 1.3.10, para finalmente aplicar el Corolario 1.3.11, obteniendo así, el resultado asintótico de normalidad. Es necesario mencionar que la demostración de la normalidad asintótica es para los estimadores de máxima log-verosimilitud “profile” del modelo. Sin embargo, posteriormente se demostrará que los estimadores de máxima log-verosimilitud parcial (2.6.12) y los estimadores de máxima log-verosimilitud “profile” bajo el estimador de Breslow para $\Lambda_0(\cdot)$ son el mismo, por tanto se obtiene la normalidad asintótica de los estimadores (2.6.12).

Con la finalidad de expresar el modelo semiparamétrico (2.2.1) como un modelo

de la forma

$$(\boldsymbol{\theta}, \Lambda_0) \mapsto P_{\boldsymbol{\theta}, \Lambda_0},$$

donde $\boldsymbol{\theta}$ es el parámetro finito dimensional y Λ_0 es el parámetro infinito dimensional, se escribe el modelo (2.2.1) como sigue

$$\lambda(t|\boldsymbol{\theta}, \mathbf{z}) = \lambda_0(t)K(\boldsymbol{\theta}, \mathbf{z}) \quad (2.6.13)$$

con $\boldsymbol{\theta} = (\boldsymbol{\beta}, a)$ los parámetros del modelo, $\lambda_0(\cdot)$ la función de riesgo base y $\mathbf{z} \in \mathbb{R}^d$ el vector de covariables. Obsérvese que $\boldsymbol{\theta} \in \mathbb{R}^d \times \mathbb{R}^+$. En este caso se tiene

$$K(\boldsymbol{\theta}, \mathbf{z}) = K(\boldsymbol{\beta}, a, \mathbf{z}) = \left(\frac{1}{1 + \exp(-\boldsymbol{\beta}'\mathbf{z})} \right)^a \text{ para } a > 0.$$

Condiciones 2.6.1. *Para demostrar las propiedades asintóticas de los estimadores del modelo anterior se asume lo siguiente:*

- i) *el tiempo observado está dentro del intervalo $[0, \tau]$,*
- ii) *el vector de covariables \mathbf{Z} es acotado,*
- iii) *el verdadero valor $\boldsymbol{\theta}_0$ del parámetro finito dimensional del modelo se encuentra en el interior de un compacto $\Theta \subset \mathbb{R}^d \times \mathbb{R}^+$ y la verdadera función de riesgo base $\Lambda_0^0(\cdot)$ está en un convexo \mathcal{B} del espacio de funciones,*
- iv) *existe una constante positiva c_1 tal que $P_0(X \geq \tau) = E_0[\mathbb{1}_{\{X \geq \tau\}}] > c_1$, donde $E_0[f]$ denota el valor esperado de f bajo la verdadera función de distribución $P_0(\cdot)$,*
- v) *la función de riesgo base acumulativa $\Lambda_0(\cdot)$ es estrictamente creciente en $[0, \tau]$ y es continuamente diferenciable, y*
- vi) *la matriz de información eficiente $\tilde{\mathbf{I}}_0$ es invertible.*

Supóngase que $\mathbf{Y}_i \sim P_{\boldsymbol{\theta}, \Lambda_0} \in \mathcal{P}$ donde \mathcal{P} es el modelo semiparamétrico dado por

$$\mathcal{P} = \{P_{\boldsymbol{\theta}, \Lambda_0} : \boldsymbol{\theta} \in \mathbb{R}^d \times \mathbb{R}^+ \text{ es el vector de parámetros y } \Lambda_0(\cdot) \text{ es la función de riesgo base acumulativa}\}$$

y la función de densidad correspondiente es $p_{\boldsymbol{\theta}, \Lambda_0}(\cdot)$.

Dado el modelo generalizado de riesgos proporcionales (2.6.13) y el mecanismo de censura por la derecha, la función de densidad para el dato \mathbf{y} es

$$\begin{aligned} p_{\boldsymbol{\theta}, \Lambda_0}(\mathbf{y}) &= [\lambda_0(x)K(\boldsymbol{\theta}, \mathbf{z})S_{T|\mathbf{Z}}(x|\mathbf{z})S_{C|\mathbf{Z}}(x|\mathbf{z})]^\delta [S_{T|\mathbf{Z}}(x|\mathbf{z})p_{C|\mathbf{Z}}(x|\mathbf{z})]^{1-\delta} p_{\mathbf{Z}}(\mathbf{z}) \\ &= [\lambda_0(x)K(\boldsymbol{\theta}, \mathbf{z}) \exp(-\Lambda_0(x)K(\boldsymbol{\theta}, \mathbf{z}))S_{C|\mathbf{Z}}(x|\mathbf{z})]^\delta \\ &\quad [\exp(-\Lambda_0(x)K(\boldsymbol{\theta}, \mathbf{z}))p_{C|\mathbf{Z}}(x|\mathbf{z})]^{1-\delta} p_{\mathbf{Z}}(\mathbf{z}), \end{aligned}$$

así, la log-verosimilitud para el dato \mathbf{y} viene dada por

$$\ell(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \log p_{\boldsymbol{\theta}, \Lambda_0}(\mathbf{y}).$$

Además, la función de log-verosimilitud para la muestra de n datos es

$$\ell_n(\boldsymbol{\theta}, \Lambda_0) = \sum_{i=1}^n \ell(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}_i)$$

y la función de log-verosimilitud “profile” para $\boldsymbol{\theta}$ es

$$p\ell_n(\boldsymbol{\theta}) = \sup_{\Lambda_0} \ell_n(\boldsymbol{\theta}, \Lambda_0). \quad (2.6.14)$$

Para verificar la condición (1.3.32) es necesario obtener la función “score” eficiente del modelo, la cual viene dada por la expresión (1.3.30). Por tanto, es necesario obtener la función “score” con respecto al parámetro $\boldsymbol{\theta}$, para ello se retoma la función de densidad $p_{\boldsymbol{\theta}, \Lambda_0}(\mathbf{y})$, nótese que ésta es proporcional a

$$[\lambda_0(x)K(\boldsymbol{\theta}, \mathbf{z})]^\delta \exp(-\Lambda_0(x)K(\boldsymbol{\theta}, \mathbf{z})),$$

de donde, el valor de log-verosimilitud para \mathbf{y} es

$$\ell(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \delta [\log \lambda_0(x) + \log K(\boldsymbol{\theta}, \mathbf{z})] - \Lambda_0(x)K(\boldsymbol{\theta}, \mathbf{z}) + cte. \quad (2.6.15)$$

La función “score” para $\boldsymbol{\theta}$ es la derivada parcial de la función de log-verosimilitud anterior con respecto a $\boldsymbol{\theta}$ y está dada por

$$\dot{\ell}(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \left(\left[\frac{\partial}{\partial \boldsymbol{\beta}} \ell(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) \right]', \frac{\partial}{\partial a} \ell(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) \right)'.$$

De la expresión (2.6.15) se tiene

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} \ell(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) &= \delta \frac{1}{K(\boldsymbol{\theta}, \mathbf{z})} \frac{\partial}{\partial \boldsymbol{\beta}} K(\boldsymbol{\theta}, \mathbf{z}) - \Lambda_0(x) \frac{\partial}{\partial \boldsymbol{\beta}} K(\boldsymbol{\theta}, \mathbf{z}) \\ &= \delta K_1(\boldsymbol{\theta}, \mathbf{z}) - \Lambda_0(x) K_1(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z}), \end{aligned} \quad (2.6.16)$$

con $K_1(\boldsymbol{\theta}, \mathbf{z}) = \frac{a\mathbf{z}}{1+\exp(\boldsymbol{\beta}'\mathbf{z})}$, ya que se tiene

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} K(\boldsymbol{\theta}, \mathbf{z}) &= a (1 + \exp(-\boldsymbol{\beta}'\mathbf{z}))^{-a-1} \exp(-\boldsymbol{\beta}'\mathbf{z})\mathbf{z} \\ &= a\mathbf{z} \frac{\exp(-\boldsymbol{\beta}'\mathbf{z})}{(1 + \exp(-\boldsymbol{\beta}'\mathbf{z}))^a (1 + \exp(-\boldsymbol{\beta}'\mathbf{z}))} \\ &= \frac{a\mathbf{z}K(\boldsymbol{\theta}, \mathbf{z})}{1 + \exp(\boldsymbol{\beta}'\mathbf{z})} \\ &= K_1(\boldsymbol{\theta}, \mathbf{z})K(\boldsymbol{\theta}, \mathbf{z}). \end{aligned}$$

Además, también de la expresión (2.6.15) se tiene

$$\begin{aligned} \frac{\partial}{\partial a} \ell(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) &= \delta \frac{1}{K(\boldsymbol{\theta}, \mathbf{z})} \frac{\partial}{\partial a} K(\boldsymbol{\theta}, \mathbf{z}) - \Lambda_0(x) \frac{\partial}{\partial a} K(\boldsymbol{\theta}, \mathbf{z}) \\ &= \delta K_2(\boldsymbol{\theta}, \mathbf{z}) - \Lambda_0(x) K_2(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z}), \end{aligned} \quad (2.6.17)$$

donde

$$\begin{aligned} \frac{\partial}{\partial a} K(\boldsymbol{\theta}, \mathbf{z}) &= K(\boldsymbol{\theta}, \mathbf{z}) \ln[K(\boldsymbol{\theta}, \mathbf{z})] \\ &= K(\boldsymbol{\theta}, \mathbf{z}) K_2(\boldsymbol{\theta}, \mathbf{z}), \end{aligned}$$

con $K_2(\boldsymbol{\theta}, \mathbf{z}) = \ln[K(\boldsymbol{\theta}, \mathbf{z})]$.

Nótese que la función “score” con respecto al parámetro $\boldsymbol{\theta}$ se puede expresar

$$\dot{\ell}(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \delta \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \Lambda_0(x) \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z}),$$

donde $\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) = (K_1(\boldsymbol{\theta}, \mathbf{z})', K_2(\boldsymbol{\theta}, \mathbf{z}))'$.

Por otro lado, para obtener la otra parte de la función “score” eficiente se define la curva

$$\Lambda_{\mathbf{t}}(x) = \int_{[0,x]} (1 + \mathbf{t}' h(s)) d\Lambda_0(s) \quad (2.6.18)$$

para $\mathbf{t} \in \mathbb{R}^{d+1}$, $\Lambda_0(\cdot)$ dada y una función acotada $h : [0, \tau] \rightarrow \mathbb{R}^{d+1}$, donde la función $h(\cdot)$ es una dirección menos favorable en el verdadero valor del parámetro Λ_0^0 . Esta curva define un submodelo de $\Lambda_0(\cdot)$. Para que esta función $\Lambda_{\mathbf{t}}(\cdot)$ sea una función de riesgo acumulativa es necesario que sea no negativa y no decreciente en $[0, \infty)$, sin embargo, como $h(\cdot)$ es una función acotada ambos requerimientos se cumplen si $\mathbf{t} \approx \mathbf{0}$.

Continuando con la construcción de la función “score” eficiente del modelo, se requiere obtener la función “score” para Λ_0 , para ello se sustituye el submodelo (2.6.18) en la función de log-verosimilitud (2.6.15) y se deriva con respecto a \mathbf{t} , así

$$\begin{aligned} \ell(\boldsymbol{\theta}, \Lambda_{\mathbf{t}})(\mathbf{y}) &= \delta [\log((1 + \mathbf{t}' h(x)) \lambda_0(x)) + \log K(\boldsymbol{\theta}, \mathbf{z})] \\ &\quad - K(\boldsymbol{\theta}, \mathbf{z}) \int_{[0,x]} (1 + \mathbf{t}' h(s)) d\Lambda_0(s) + cte, \end{aligned}$$

y

$$\frac{\partial}{\partial \mathbf{t}} \ell(\boldsymbol{\theta}, \Lambda_{\mathbf{t}})(\mathbf{y}) = \delta \frac{h(x) \lambda_0(x)}{(1 + \mathbf{t}' h(x)) \lambda_0(x)} - K(\boldsymbol{\theta}, \mathbf{z}) \int_{[0,x]} h(s) d\Lambda_0(s).$$

Como el submodelo $\Lambda_{\mathbf{t}}(\cdot)$ pasa a través de $\Lambda_0(\cdot)$ en $\mathbf{t} = \mathbf{0}$, se evalúa esta derivada

en $\mathbf{t} = \mathbf{0}$, así

$$\begin{aligned} \left. \frac{\partial}{\partial \mathbf{t}} \ell(\boldsymbol{\theta}, \Lambda_{\mathbf{t}})(\mathbf{y}) \right|_{\mathbf{t}=\mathbf{0}} &= \delta \frac{h(x)\lambda_0(x)}{(1 + \mathbf{t}'h(x))\lambda_0(x)} \Big|_{\mathbf{t}=\mathbf{0}} - K(\boldsymbol{\theta}, \mathbf{z}) \int_{[0,x]} h(s) d\Lambda_0(s) \\ &= \delta h(x) - K(\boldsymbol{\theta}, \mathbf{z}) \int_{[0,x]} h(s) d\Lambda_0(s) \\ &= A_{\boldsymbol{\theta}, \Lambda_0} h(\mathbf{y}). \end{aligned}$$

Dado que $\Pi_{\boldsymbol{\theta}, \Lambda_0}$ es la proyección ortogonal sobre la clausura del espacio lineal generado por $\dot{\mathcal{P}}_{\boldsymbol{\theta}, \Lambda_0}$ entonces, al igual que en (1.3.30), la función “score” eficiente para $\boldsymbol{\theta}$ estaría dada por:

$$\tilde{\ell}(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \dot{\ell}(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) - A_{\boldsymbol{\theta}, \Lambda_0} \left(\frac{A_{\boldsymbol{\theta}, \Lambda_0}^* \dot{\ell}(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y})}{A_{\boldsymbol{\theta}, \Lambda_0}^* A_{\boldsymbol{\theta}, \Lambda_0}(x)} \right). \quad (2.6.19)$$

donde $A_{\boldsymbol{\theta}, \Lambda_0}^*$ es el operador adjunto caracterizado por

$$P_{\boldsymbol{\theta}, \Lambda_0} [(A_{\boldsymbol{\theta}, \Lambda_0} h)g] = \langle A_{\boldsymbol{\theta}, \Lambda_0} h, g \rangle_{P_{\boldsymbol{\theta}, \Lambda_0}} = \langle h, A_{\boldsymbol{\theta}, \Lambda_0}^* g \rangle_{\Lambda_0} = \Lambda_0 [h(A_{\boldsymbol{\theta}, \Lambda_0}^* g)]. \quad (2.6.20)$$

Para obtener el segundo sumando de la expresión (2.6.19), previamente se deben obtener las expresiones de $A_{\boldsymbol{\theta}, \Lambda_0}^* \dot{\ell}(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y})$ y $A_{\boldsymbol{\theta}, \Lambda_0}^* A_{\boldsymbol{\theta}, \Lambda_0}(x)$. Para ello se usan las siguientes igualdades, deducidas a partir de (2.6.20).

$$\begin{aligned} P_{\boldsymbol{\theta}, \Lambda_0} [(A_{\boldsymbol{\theta}, \Lambda_0} g)(A_{\boldsymbol{\theta}, \Lambda_0} h)] &= \langle A_{\boldsymbol{\theta}, \Lambda_0} g, A_{\boldsymbol{\theta}, \Lambda_0} h \rangle_{P_{\boldsymbol{\theta}, \Lambda_0}} \\ &= \langle g, A_{\boldsymbol{\theta}, \Lambda_0}^* A_{\boldsymbol{\theta}, \Lambda_0} h \rangle_{\Lambda_0} \\ &= \Lambda_0 [g(A_{\boldsymbol{\theta}, \Lambda_0}^* A_{\boldsymbol{\theta}, \Lambda_0} h)] \end{aligned} \quad (2.6.21)$$

y

$$\begin{aligned} \Lambda_0 [(A_{\boldsymbol{\theta}, \Lambda_0}^* \dot{\ell}(\boldsymbol{\theta}, \Lambda_0))h] &= \Lambda_0 [h(A_{\boldsymbol{\theta}, \Lambda_0}^* \dot{\ell}(\boldsymbol{\theta}, \Lambda_0))] \\ &= \langle h, A_{\boldsymbol{\theta}, \Lambda_0}^* \dot{\ell}(\boldsymbol{\theta}, \Lambda_0) \rangle_{\Lambda_0} \\ &= \langle A_{\boldsymbol{\theta}, \Lambda_0} h, \dot{\ell}(\boldsymbol{\theta}, \Lambda_0) \rangle_{P_{\boldsymbol{\theta}, \Lambda_0}} \\ &= P_{\boldsymbol{\theta}, \Lambda_0} [(A_{\boldsymbol{\theta}, \Lambda_0} h) \dot{\ell}(\boldsymbol{\theta}, \Lambda_0)] \\ &= P_{\boldsymbol{\theta}, \Lambda_0} [\dot{\ell}(\boldsymbol{\theta}, \Lambda_0)(A_{\boldsymbol{\theta}, \Lambda_0} h)]. \end{aligned} \quad (2.6.22)$$

Entonces, para obtener la expresión $A_{\boldsymbol{\theta}, \Lambda_0}^* A_{\boldsymbol{\theta}, \Lambda_0}(x)$, en primer lugar, se calcula el producto $(A_{\boldsymbol{\theta}, \Lambda_0} g)(A_{\boldsymbol{\theta}, \Lambda_0} h)$ que viene dado por

$$\begin{aligned}
 & (A_{\boldsymbol{\theta}, \Lambda_0} g(\mathbf{y}))'(A_{\boldsymbol{\theta}, \Lambda_0} h(\mathbf{y})) \\
 &= \left(\delta g(x) - K(\boldsymbol{\theta}, \mathbf{z}) \int_{[0, x]} g(s) d\Lambda_0(s) \right)' \left(\delta h(x) - K(\boldsymbol{\theta}, \mathbf{z}) \int_{[0, x]} h(s) d\Lambda_0(s) \right) \\
 &= \delta g(x)' h(x) - \delta g(x)' K(\boldsymbol{\theta}, \mathbf{z}) \int_{[0, x]} h(s) d\Lambda_0(s) - \delta h(x)' K(\boldsymbol{\theta}, \mathbf{z}) \int_{[0, x]} g(s) d\Lambda_0(s) \\
 &\quad + (K(\boldsymbol{\theta}, \mathbf{z}))^2 \left(\int_{[0, x]} g(s) d\Lambda_0(s) \right)' \int_{[0, x]} h(s) d\Lambda_0(s) \\
 &= \delta g(x)' h(x) - \delta K(\boldsymbol{\theta}, \mathbf{z}) \left[g(x)' \int_{[0, x]} h(s) d\Lambda_0(s) + h(x)' \int_{[0, x]} g(s) d\Lambda_0(s) \right] \\
 &\quad + (K(\boldsymbol{\theta}, \mathbf{z}))^2 \left(\int_{[0, x]} g(s) d\Lambda_0(s) \right)' \int_{[0, x]} h(s) d\Lambda_0(s).
 \end{aligned}$$

Se define

$$a(x) = g(x)' \int_{[0, x]} h(s) d\Lambda_0(s) + h(x)' \int_{[0, x]} g(s) d\Lambda_0(s)$$

e integrando por partes como en el Apéndice II se tiene que

$$(A_{\boldsymbol{\theta}, \Lambda_0} g(\mathbf{y}))'(A_{\boldsymbol{\theta}, \Lambda_0} h(\mathbf{y})) = \delta g(x)' h(x) - \delta K(\boldsymbol{\theta}, \mathbf{z}) a(x) + (K(\boldsymbol{\theta}, \mathbf{z}))^2 \int_{[0, x]} a(s) d\Lambda_0(s).$$

Después, se calcula $P_{\boldsymbol{\theta}, \Lambda_0}$ del segundo término que viene dado por

$$\begin{aligned}
 & P_{\boldsymbol{\theta}, \Lambda_0} [-\delta K(\boldsymbol{\theta}, \mathbf{z}) a(x)] \\
 &= - \int_0^\tau \int_{\mathbb{R}^d} K(\boldsymbol{\theta}, \mathbf{z}) a(x) \lambda_0(x) K(\boldsymbol{\theta}, \mathbf{z}) S_{T|\mathbf{Z}}(x|\mathbf{z}) S_{C|\mathbf{Z}}(x|\mathbf{z}) p_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} dx \\
 &= - \int_0^\tau a(x) \int_{\mathbb{R}^d} (K(\boldsymbol{\theta}, \mathbf{z}))^2 S_{T|\mathbf{Z}}(x|\mathbf{z}) S_{C|\mathbf{Z}}(x|\mathbf{z}) p_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} \lambda_0(x) dx \\
 &= - \int_0^\tau a(x) \int_{\mathbb{R}^d} (K(\boldsymbol{\theta}, \mathbf{z}))^2 P[T \wedge C > x|\mathbf{z}] p_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} d\Lambda_0(x) \\
 &= - \int_0^\tau a(x) \int_{\mathbb{R}^d} (K(\boldsymbol{\theta}, \mathbf{z}))^2 E[\mathbb{1}_{\{X \geq x\}}|\mathbf{z}] p_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} d\Lambda_0(x) \\
 &= - \int_0^\tau a(x) E[(K(\boldsymbol{\theta}, \mathbf{z}))^2 \mathbb{1}_{\{X \geq x\}}] d\Lambda_0(x).
 \end{aligned}$$

Por tanto, se cancelan el segundo y tercer término de la integral y sólo queda el primer término de dicha integral, de donde se tiene

$$\begin{aligned}
 P_{\boldsymbol{\theta}, \Lambda_0}[\delta g(x)'h(x)] &= \int_0^\tau \int_{\mathbb{R}^d} g(x)'h(x)K(\boldsymbol{\theta}, \mathbf{z})\lambda_0(x)S_{T|\mathbf{Z}}(x|\mathbf{z})S_{C|\mathbf{Z}}(x|\mathbf{z})p_{\mathbf{Z}}(\mathbf{z})d\mathbf{z}dx \\
 &= \int_0^\tau g(x)'h(x) \int_{\mathbb{R}^d} K(\boldsymbol{\theta}, \mathbf{z})S_{T|\mathbf{Z}}(x|\mathbf{z})S_{C|\mathbf{Z}}(x|\mathbf{z})p_{\mathbf{Z}}(\mathbf{z})d\mathbf{z}\lambda_0(x)dx \\
 &= \int_0^\tau g(x)'h(x) \left[\int_{\mathbb{R}^d} K(\boldsymbol{\theta}, \mathbf{z})P[T \wedge C > x|\mathbf{z}]p_{\mathbf{Z}}(\mathbf{z})d\mathbf{z} \right] d\Lambda_0(x) \\
 &= \int_0^\tau g(x)'h(x) \left[\int_{\mathbb{R}^d} K(\boldsymbol{\theta}, \mathbf{z})E[\mathbb{1}_{\{X \geq x\}}|\mathbf{z}]p_{\mathbf{Z}}(\mathbf{z})d\mathbf{z} \right] d\Lambda_0(x) \\
 &= \int_0^\tau g(x)'E[K(\boldsymbol{\theta}, \mathbf{z})\mathbb{1}_{\{X \geq x\}}]h(x)d\Lambda_0(x) \\
 &= \Lambda_0 \left[g'E[K(\boldsymbol{\theta}, \mathbf{z})\mathbb{1}_{\{X \geq x\}}]h \right].
 \end{aligned}$$

Finalmente, por la ecuación (2.6.21) se tiene

$$A_{\boldsymbol{\theta}, \Lambda_0}^* A_{\boldsymbol{\theta}, \Lambda_0} h(\mathbf{y}) = E[K(\boldsymbol{\theta}, \mathbf{z})\mathbb{1}_{\{X \geq x\}}]h(\mathbf{y}).$$

Análogamente, tomando el producto $\dot{\ell}(\boldsymbol{\theta}, \Lambda_0)A_{\boldsymbol{\theta}, \Lambda_0}h$, integrando y usando la expresión (2.6.22) se tiene

$$A_{\boldsymbol{\theta}, \Lambda_0}^* \dot{\ell}(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = E[\tilde{K}(\boldsymbol{\theta}, \mathbf{z})K(\boldsymbol{\theta}, \mathbf{z})\mathbb{1}_{\{X \geq x\}}].$$

Por otro lado, si se definen las siguientes funciones

$$M_{\boldsymbol{\theta}, \Lambda_0}^0(x) = E_{\boldsymbol{\theta}, \Lambda_0} [K(\boldsymbol{\theta}, \mathbf{z})\mathbb{1}_{\{X \geq x\}}]$$

y

$$M_{\boldsymbol{\theta}, \Lambda_0}^1(x) = E_{\boldsymbol{\theta}, \Lambda_0} [\tilde{K}(\boldsymbol{\theta}, \mathbf{z})K(\boldsymbol{\theta}, \mathbf{z})\mathbb{1}_{\{X \geq x\}}]$$

y tomando de nuevo la expresión (2.6.19) se tiene que la función “score” eficiente

para $\boldsymbol{\theta}$ es

$$\tilde{\ell}(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \dot{\ell}(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) - A_{\boldsymbol{\theta}, \Lambda_0} \left(\frac{A_{\boldsymbol{\theta}, \Lambda_0}^* \dot{\ell}(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y})}{A_{\boldsymbol{\theta}, \Lambda_0}^* A_{\boldsymbol{\theta}, \Lambda_0}(x)} \right) \quad (2.6.23)$$

$$\begin{aligned} &= \delta \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \Lambda_0(x) \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z}) - \delta \frac{M_{\boldsymbol{\theta}, \Lambda_0}^1(x)}{M_{\boldsymbol{\theta}, \Lambda_0}^0(x)} \\ &\quad + K(\boldsymbol{\theta}, \mathbf{z}) \int_{[0, x]} \frac{M_{\boldsymbol{\theta}, \Lambda_0}^1(s)}{M_{\boldsymbol{\theta}, \Lambda_0}^0(s)} d\Lambda_0(s) \\ &= \delta \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{M_{\boldsymbol{\theta}, \Lambda_0}^1(x)}{M_{\boldsymbol{\theta}, \Lambda_0}^0(x)} \right) \\ &\quad - K(\boldsymbol{\theta}, \mathbf{z}) \left(\Lambda_0(x) \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \int_{[0, x]} \frac{M_{\boldsymbol{\theta}, \Lambda_0}^1(s)}{M_{\boldsymbol{\theta}, \Lambda_0}^0(s)} d\Lambda_0(s) \right) \\ &= \delta \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{M_{\boldsymbol{\theta}, \Lambda_0}^1(x)}{M_{\boldsymbol{\theta}, \Lambda_0}^0(x)} \right) \\ &\quad - K(\boldsymbol{\theta}, \mathbf{z}) \int_{[0, x]} \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{M_{\boldsymbol{\theta}, \Lambda_0}^1(s)}{M_{\boldsymbol{\theta}, \Lambda_0}^0(s)} \right) d\Lambda_0(s). \end{aligned} \quad (2.6.24)$$

Por tanto, la expresión (2.6.24) es la función “score” eficiente del modelo logit generalizado con datos con censura por la derecha.

Por otra parte, supóngase que $\boldsymbol{\theta}_0$ y $\Lambda_0^0(\cdot)$ son los verdaderos parámetros del modelo semiparamétrico. Entonces se puede tomar la función

$$h_0(x) = \frac{M_{\boldsymbol{\theta}_0, \Lambda_0^0}^1(x)}{M_{\boldsymbol{\theta}_0, \Lambda_0^0}^0(x)}$$

como la dirección menos favorable en el submodelo

$$\Lambda_{\mathbf{t}}(x) = \int_{[0, x]} (1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)) d\Lambda_0(s),$$

de donde se puede ver que para $\mathbf{t} = \boldsymbol{\theta}$ se obtiene $\Lambda_0(\cdot)$, es decir, se cumple la condición (1.3.31) del Teorema 1.3.10.

La función de log-verosimilitud para este submodelo es

$$\ell(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \log p_{\mathbf{t}, \Lambda_{\mathbf{t}}}(\mathbf{y}).$$

Además, se tiene que la función “score” del submodelo es

$$\dot{\ell}(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \frac{\partial}{\partial \mathbf{t}} \ell(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0)(\mathbf{y}),$$

evaluando la expresión anterior en $(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0, \Lambda_0^0)$ se puede obtener

$$\dot{\ell}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0, \Lambda_0^0)(\mathbf{y}) = \dot{\ell}(\boldsymbol{\theta}_0, \Lambda_0^0)(\mathbf{y}) - A_{\boldsymbol{\theta}_0, \Lambda_0^0} h_0(\mathbf{y})$$

y tomando la función “score” eficiente (2.6.24) evaluada en $(\boldsymbol{\theta}_0, \Lambda_0^0)$ es fácil ver que

$$\dot{\ell}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0, \Lambda_0^0)(\mathbf{y}) = \tilde{\ell}(\boldsymbol{\theta}_0, \Lambda_0^0)(\mathbf{y}).$$

Por tanto, se cumple la condición (1.3.32) del Teorema 1.3.10.

Ahora, antes de verificar las condiciones (1.3.33) y (1.3.34), se verificará si se cumplen las hipótesis del Teorema 1.3.10 para ello es necesario demostrar los siguientes lemas.

Lema 2.6.2. *La clase de funciones*

$$\{\dot{\ell}(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0) : (\mathbf{t}, \boldsymbol{\theta}, \Lambda_0) \in V\}$$

es Donsker con función envolvente cuadrado integrable, donde V es una vecindad de $(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0, \Lambda_0^0)$.

Demostración. Sea \mathcal{F} el conjunto de funciones de distribución continuas. Para $\rho > 0$ se define

$$C_\rho = \{P \in \mathcal{F} : \|P - P_0\|_\infty \leq \rho\},$$

donde $P_0 = P_{\boldsymbol{\theta}_0, \Lambda_0^0}$, es decir, la verdadera función de distribución.

Por otro lado, como el vector \mathbf{z} es acotado. La clase $\{\boldsymbol{\beta}'\mathbf{z} : \boldsymbol{\beta} \in \Theta_\beta\}$ es Donsker, con Θ_β tal que $\Theta = \Theta_\beta \times \Theta_a$. Dado que la función $\exp(\boldsymbol{\beta}'\mathbf{z})$ es diferenciable y que las derivadas son acotadas, se tiene que la clase $\{\exp(\boldsymbol{\beta}'\mathbf{z}) : \boldsymbol{\beta} \in \Theta_\beta\}$ es Donsker. Además como $1 + \exp(\boldsymbol{\beta}'\mathbf{z}) > 0$ entonces

$$\left\{ K(\boldsymbol{\theta}, \mathbf{z}) = \left(\frac{1}{1 + \exp(-\boldsymbol{\beta}'\mathbf{z})} \right)^a : \boldsymbol{\theta} \in \Theta \right\}, \quad (2.6.25)$$

y

$$\left\{ K_1(\boldsymbol{\theta}, \mathbf{z}) = \frac{a\mathbf{z}}{1 + \exp(\boldsymbol{\beta}'\mathbf{z})} : \boldsymbol{\theta} \in \Theta \right\}$$

son Donsker. Además, como la función $f(x) = \ln(x)$ con dominio en $[c, \infty)$ donde $c > 0$ es Lipschitz se obtiene que la clase

$$\{K_2(\boldsymbol{\theta}, \mathbf{z}) = \ln[K(\boldsymbol{\theta}, \mathbf{z})] : \boldsymbol{\theta} \in \Theta\}$$

es Donsker. Entonces,

$$\left\{ \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) : \boldsymbol{\theta} \in \Theta \right\}$$

también es Donsker.

Además, se tiene que la clase $\{\mathbb{1}_{\{X \geq x\}} : x \in [0, \tau]\}$ es Donsker, por tanto, las clases $\{K(\boldsymbol{\theta}, \mathbf{z})\mathbb{1}_{\{X \geq x\}} : x \in [0, \tau], \boldsymbol{\theta} \in \Theta\}$ y $\{\tilde{K}(\boldsymbol{\theta}, \mathbf{z})K(\boldsymbol{\theta}, \mathbf{z})\mathbb{1}_{\{X \geq x\}} : x \in [0, \tau], \boldsymbol{\theta} \in \Theta\}$ son Donsker. Para $P \in C_\rho$, como la función $P \rightarrow E_P[f]$ es Lipschitz, las clases

$$\{E_P [K(\boldsymbol{\theta}, \mathbf{z})\mathbb{1}_{\{X \geq x\}}] : x \in [0, \tau], \boldsymbol{\theta} \in \Theta, P \in C_\rho\}$$

y

$$\left\{ E_P \left[\tilde{K}(\boldsymbol{\theta}, \mathbf{z})K(\boldsymbol{\theta}, \mathbf{z})\mathbb{1}_{\{X \geq x\}} \right] : x \in [0, \tau], \boldsymbol{\theta} \in \Theta, P \in C_\rho \right\}$$

son Donsker.

Por otro lado, como \mathbf{z} es acotado y $\boldsymbol{\theta} \in \Theta$, existen m, M tal que

$$0 < m < K(\boldsymbol{\theta}, \mathbf{z}) < M < \infty. \quad (2.6.26)$$

Como la función $P \rightarrow E_P[f]$ es continua, existe $\rho_1 > 0$ tal que para todo $P \in C_\rho$ se tiene

$$E_P [\mathbb{1}_{\{X \geq x\}}] \geq \rho_1 > 0. \quad (2.6.27)$$

De las expresiones (2.6.26) y (2.6.27) se obtiene

$$0 < \rho_1 m \leq m E_P [\mathbb{1}_{\{X \geq x\}}] \leq E_P [K(\boldsymbol{\theta}, \mathbf{z})\mathbb{1}_{\{X \geq x\}}] \leq M E_P [\mathbb{1}_{\{X \geq x\}}] < \infty$$

entonces la clase

$$\left\{ \frac{1}{E_P [K(\boldsymbol{\theta}, \mathbf{z}) \mathbb{1}_{\{X \geq x\}}]} : x \in [0, \tau], \boldsymbol{\theta} \in \Theta, P \in C_\rho \right\}$$

es Donsker y por consiguiente la clase

$$\left\{ \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{E_P [\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z}) \mathbb{1}_{\{X \geq x\}}]}{E_P [K(\boldsymbol{\theta}, \mathbf{z}) \mathbb{1}_{\{X \geq x\}}]} : x \in [0, \tau], \boldsymbol{\theta} \in \Theta, P \in C_\rho \right\} \quad (2.6.28)$$

lo es también.

Como se sabe la función $\Lambda_0 \rightarrow \int f d\Lambda_0$ es Lipschitz entonces la clase

$$\left\{ \int_{[0, x]} \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{E_P [\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z}) \mathbb{1}_{\{X \geq x\}}]}{E_P [K(\boldsymbol{\theta}, \mathbf{z}) \mathbb{1}_{\{X \geq x\}}]} \right) d\Lambda_0 : x \in [0, \tau], \boldsymbol{\theta} \in \Theta, P \in C_\rho \right\} \quad (2.6.29)$$

es Donsker. Finalmente, como consecuencia de (2.6.25), (2.6.28) y (2.6.29) se obtiene que

$$\{\tilde{\ell}(\boldsymbol{\theta}, \Lambda_0) : x \in [0, \tau], \boldsymbol{\theta} \in \Theta, P \in C_\rho\}$$

es Donsker, y por tanto, con función envolvente cuadrado integrable. \square

Lema 2.6.3. *La clase de funciones*

$$\{\ddot{\ell}(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0) : (\mathbf{t}, \boldsymbol{\theta}, \Lambda_0) \in V\}$$

es Glivenko–Cantelli con función envolvente integrable, con V una vecindad de $(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0, \Lambda_0^0)$.

Demostración. La demostración es análoga a la del Lema 2.6.2. \square

Los lemas anteriores verifican las hipótesis del Teorema 1.3.10 sobre la primera y segunda derivadas de la función de log-verosimilitud del submodelo menos favorable del modelo logit generalizado con datos con censura por la derecha. Sin embargo, aún falta verificar que se cumplen las condiciones (1.3.33) y (1.3.34) para

poder aplicar dicho teorema. Esto es, es necesario demostrar la consistencia de los estimadores del modelo. Para ello, se define el estimador de la función $\Lambda_0(\cdot)$ como

$$\hat{\Lambda}_{\boldsymbol{\theta}}(x) = \int_{[0,x]} \frac{dN(s)}{E_{P_Y}[\mathbb{1}_{\{X \geq s\}} K(\boldsymbol{\theta}, \mathbf{z})]} = P_Y \left[\frac{N(x)}{P_Y[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \right], \quad (2.6.30)$$

con $N(x) = \mathbb{1}_{\{X \leq x, \delta=1\}}$. Obsérvese que el estimador anterior depende del parámetro $\boldsymbol{\theta}$ y de la función de distribución $P_Y(\cdot)$. En particular, si se considera la función de distribución empírica se tiene

$$\begin{aligned} \hat{\Lambda}_{\boldsymbol{\theta}}(x) &= \mathbb{P}_n \left[\frac{N(x)}{\mathbb{P}_n[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \right] \\ &= \frac{1}{n} \sum_{i=1}^n \frac{N_i(x)}{\mathbb{P}_n[\mathbb{1}_{\{X \geq x_i\}} K(\boldsymbol{\theta}, \mathbf{z})]}. \end{aligned} \quad (2.6.31)$$

con $N_i(x) = \mathbb{1}_{\{X_i \leq x, \delta_i=1\}}$. El estimador anterior es el equivalente al estimador de Breslow en el modelo de Cox. Más adelante, en la Sección 2.6.5, se estudiará otra forma de obtener dicho estimador.

También, nótese que si en el estimador (2.6.30) se toma el verdadero valor del parámetro $\boldsymbol{\theta}_0$ y la verdadera función de distribución $P_0(\cdot)$ se tiene

$$\begin{aligned} \hat{\Lambda}_{\boldsymbol{\theta}_0}(x) &= P_0 \left[\frac{N(x)}{E_0[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}_0, \mathbf{z})]} \right] \\ &= \int_{\mathbb{R}^d} \int_0^\tau \frac{\mathbb{1}_{\{X \leq s\}} \lambda_0^0(s) K(\boldsymbol{\theta}_0, \mathbf{z}) S_{T|\mathbf{Z}}(s|\mathbf{z}) S_{C|\mathbf{Z}}(s|\mathbf{z}) p(\mathbf{z}) ds d\mathbf{z}}{E_0[\mathbb{1}_{\{X \geq s\}} K(\boldsymbol{\theta}_0, \mathbf{z})]} \\ &= \int_{[0,x]} \frac{\lambda_0^0(s) [\int_{\mathbb{R}^d} \mathbb{1}_{\{X \geq s\}} K(\boldsymbol{\theta}_0, \mathbf{z}) S_{T|\mathbf{Z}}(s|\mathbf{z}) S_{C|\mathbf{Z}}(s|\mathbf{z}) p(\mathbf{z}) d\mathbf{z}] ds}{E_0[\mathbb{1}_{\{X \geq s\}} K(\boldsymbol{\theta}_0, \mathbf{z})]} \\ &= \int_{[0,x]} \lambda_0^0(s) \frac{E_0[\mathbb{1}_{\{X \geq s\}} K(\boldsymbol{\theta}_0, \mathbf{z})]}{E_0[\mathbb{1}_{\{X \geq s\}} K(\boldsymbol{\theta}_0, \mathbf{z})]} ds \\ &= \int_{[0,x]} \lambda_0^0(s) ds \\ &= \Lambda_0^0(x). \end{aligned}$$

Para demostrar la consistencia de los estimadores del modelo generalizado de riesgos proporcionales se sigue el procedimiento de Murphy (1994) y se usarán los siguientes lemas.

Lema 2.6.4. *El estimador (2.6.31) en $\boldsymbol{\theta}_0$ es consistente, es decir,*

$$\sup_{x \in [0, \tau]} |\hat{\Lambda}_{\boldsymbol{\theta}_0}(x) - \Lambda_0^0(x)| \xrightarrow{c.s.} 0,$$

cuando $n \rightarrow \infty$.

Demostración. Por argumentos similares a los del Lema 2.6.2 se puede demostrar que las clases de funciones

$$\{\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z}) : x \in [0, \tau], \boldsymbol{\theta} \in \Theta\} \quad (2.6.32)$$

y

$$\left\{ \frac{N(x)}{P[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})]} : x \in [0, \tau], \boldsymbol{\theta} \in \Theta, P \in \mathcal{C}_\rho \right\}$$

son Donsker, y por tanto Glivenko–Cantelli, donde

$$\mathcal{C}_\rho = \{P \in \mathcal{F} : \|P - P_0\| \leq \rho\},$$

con \mathcal{F} el conjunto de distribuciones continuas y $\rho > 0$.

En particular, la clase

$$\left\{ \frac{N(x)}{P_0[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})]} : x \in [0, \tau], \boldsymbol{\theta} \in \Theta \right\} \quad (2.6.33)$$

es Donsker y por tanto Glivenko–Cantelli.

También se tiene

$$\begin{aligned}
& \sup_{x \in [0, \tau]} |\hat{\Lambda}_{\theta_0}(x) - \Lambda_0^0(x)| \\
&= \sup_{x \in [0, \tau]} \left| \mathbb{P}_n \left[\frac{N(x)}{\mathbb{P}_n[\mathbb{1}_{\{X \geq x\}} K(\theta_0, \mathbf{z})]} \right] - P_0 \left[\frac{N(x)}{P_0[\mathbb{1}_{\{X \geq x\}} K(\theta_0, \mathbf{z})]} \right] \right| \\
&\leq \sup_{x \in [0, \tau]} \left| \mathbb{P}_n \left[\frac{N(x)}{\mathbb{P}_n[\mathbb{1}_{\{X \geq x\}} K(\theta_0, \mathbf{z})]} \right] - \mathbb{P}_n \left[\frac{N(x)}{P_0[\mathbb{1}_{\{X \geq x\}} K(\theta_0, \mathbf{z})]} \right] \right| \\
&\quad + \sup_{x \in [0, \tau]} \left| \mathbb{P}_n \left[\frac{N(x)}{P_0[\mathbb{1}_{\{X \geq x\}} K(\theta_0, \mathbf{z})]} \right] - P_0 \left[\frac{N(x)}{P_0[\mathbb{1}_{\{X \geq x\}} K(\theta_0, \mathbf{z})]} \right] \right| \\
&= \sup_{x \in [0, \tau]} \left| \frac{1}{n} \sum_{i=1}^n N_i(x) \left[\frac{1}{\mathbb{P}_n[\mathbb{1}_{\{X_i \geq x\}} K(\theta_0, \mathbf{z})]} - \frac{1}{P_0[\mathbb{1}_{\{X_i \geq x\}} K(\theta_0, \mathbf{z})]} \right] \right| \\
&\quad + \sup_{x \in [0, \tau]} \left| (\mathbb{P}_n - P_0) \left[\frac{N(x)}{P_0[\mathbb{1}_{\{X \geq x\}} K(\theta_0, \mathbf{z})]} \right] \right| \\
&\leq \sup_{x \in [0, \tau]} \left| \frac{1}{\mathbb{P}_n[\mathbb{1}_{\{X_i \geq x\}} K(\theta_0, \mathbf{z})]} - \frac{1}{P_0[\mathbb{1}_{\{X_i \geq x\}} K(\theta_0, \mathbf{z})]} \right| \\
&\quad + \sup_{x \in [0, \tau]} \left| (\mathbb{P}_n - P_0) \left[\frac{N(x)}{P_0[\mathbb{1}_{\{X \geq x\}} K(\theta_0, \mathbf{z})]} \right] \right|.
\end{aligned}$$

Como la clase dada en (2.6.32) es Glivenko–Cantelli, se tiene que

$$\sup_{x \in [0, \tau]} |(\mathbb{P}_n - P_0) [\mathbb{1}_{\{X_i \geq x\}} K(\theta_0, \mathbf{z})]| \xrightarrow{c.s.} 0,$$

además, como $P_0 [\mathbb{1}_{\{X_i \geq x\}} K(\theta_0, \mathbf{z})] > 0$ para $x \in [0, \tau]$, entonces

$$\sup_{x \in [0, \tau]} \left| \frac{1}{\mathbb{P}_n[\mathbb{1}_{\{X_i \geq x\}} K(\theta_0, \mathbf{z})]} - \frac{1}{P_0[\mathbb{1}_{\{X_i \geq x\}} K(\theta_0, \mathbf{z})]} \right| \xrightarrow{c.s.} 0.$$

También, como la clase dada en (2.6.33) es Glivenko–Cantelli se tiene que

$$\sup_{x \in [0, \tau]} \left| (\mathbb{P}_n - P_0) \left[\frac{N(x)}{P_0[\mathbb{1}_{\{X \geq x\}} K(\theta_0, \mathbf{z})]} \right] \right| \xrightarrow{c.s.} 0.$$

Por tanto, se obtiene el resultado del lema. □

Lema 2.6.5. *Bajo las Condiciones 2.6.1, se obtiene que $\limsup \hat{\Lambda}_{\boldsymbol{\theta}}(\tau) < \infty$ casi seguramente.*

Demostración. Se toma $x \in [0, \tau]$, bajo la suposición de que el vector de covariables \mathbf{z} es acotado, se tiene que

$$\|\mathbf{z}\| < c,$$

para alguna constante $0 < c < \infty$.

Se define la constante

$$c_2 = \min_{\|\mathbf{z}\| < c} K(\boldsymbol{\theta}, \mathbf{z})$$

para $\boldsymbol{\theta}$ dado.

Entonces,

$$\mathbb{P}_n[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})] \geq c_2 \mathbb{P}_n[\mathbb{1}_{\{X \geq x\}}],$$

y por la ley de los grandes números, se obtiene que, casi seguramente

$$\mathbb{P}_n[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})] \geq c_2 P_0[\mathbb{1}_{\{X \geq x\}}] + o_{P_0}(1).$$

Se asume que $P_0[\mathbb{1}_{\{X \geq x\}}] > 0$ para cada $x \in [0, \tau]$, de donde

$$\mathbb{P}_n[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})] > 0$$

casi seguramente cuando $n \rightarrow \infty$.

Esto es, los saltos de $\hat{\Lambda}_{\boldsymbol{\theta}}$ en τ están acotados por $1/c_2$, entonces

$$0 \leq \hat{\Lambda}_{\boldsymbol{\theta}}(\tau) \leq O(1) \mathbb{P}_n N(\tau) / c_2$$

casi seguro cuando $n \rightarrow \infty$, con $N(x) = \mathbb{1}_{\{X \leq x, \delta=1\}}$. □

Ahora, se puede demostrar el siguiente teorema de consistencia, donde $\tilde{\boldsymbol{\theta}}_n$ es el estimador de máxima log-verosimilitud “profile”.

Teorema 2.6.6. *Bajo las Condiciones 2.6.1, los estimadores del modelo generalizado de riesgos proporcionales son consistentes, esto es*

$$\sup_{x \in [0, \tau]} \left| \hat{\Lambda}_{\tilde{\theta}_n}(x) - \Lambda_0^0(x) \right| \text{ y } \|\tilde{\theta}_n - \theta_0\|$$

convergen a 0 casi seguro cuando $n \rightarrow \infty$.

Demostración. Por el Lema 2.6.5 y el Teorema de selección de Helly, se tiene que a lo largo de una subsucesión

$$\hat{\Lambda}_{\tilde{\theta}_n}(x) \rightarrow \Lambda^*(x) \text{ para todo } x \in [0, \tau],$$

donde $\Lambda^*(\cdot)$ es una función continua no creciente y $\tilde{\theta}_n \rightarrow \theta^*$ con $\theta^* \in \Theta$.

Por otra parte, puesto que $\mathbb{P}_n[\mathbb{1}_{\{X \geq x\}} K(\theta, \mathbf{z})] > \epsilon > 0$, se tiene que $\hat{\Lambda}_{\tilde{\theta}_n}(\cdot)$ es absolutamente continua con respecto a $\hat{\Lambda}_{\theta_0}(\cdot)$ y $\frac{d\hat{\Lambda}_{\tilde{\theta}_n}(x)}{d\hat{\Lambda}_{\theta_0}(x)}$ converge a una función medible acotada $\zeta(x)$, esto es

$$\Lambda^*(x) = \int_{[0, x]} \zeta(s) d\Lambda_0^0(s).$$

Por tanto, $\Lambda^*(\cdot)$ es absolutamente continua con respecto a $\Lambda_0^0(\cdot)$, su derivada se denota por $\lambda^*(\cdot)$, y mas aún $\zeta(x) = \frac{\lambda^*(x)}{\lambda_0^0(x)}$.

Además, como $(\tilde{\theta}_n, \hat{\Lambda}_{\tilde{\theta}_n})$ maximiza $\ell_n(\theta, \Lambda_0)$ se tiene

$$0 \leq \frac{1}{n} \sum_{i=1}^n \left[\ell(\tilde{\theta}_n, \hat{\Lambda}_{\tilde{\theta}_n})(\mathbf{y}) - \ell(\theta_0, \hat{\Lambda}_{\theta_0})(\mathbf{y}) \right],$$

si $n \rightarrow \infty$ se tiene

$$0 \leq P_0[\ell(\theta^*, \Lambda^*)(\mathbf{y}) - \ell(\theta_0, \Lambda_0^0)(\mathbf{y})].$$

Por el Teorema Glivenko–Cantelli y como $\frac{d\hat{\Lambda}_{\tilde{\theta}_n}(x)}{d\hat{\Lambda}_{\theta_0}(x)}$ converge uniformemente a $\frac{\lambda^*(x)}{\lambda_0^0(x)}$, se concluye que la información de Kullback–Leibler entre la densidad dada por los parámetros θ^* y $\Lambda^*(\cdot)$ y los parámetros θ_0 y $\Lambda_0^0(\cdot)$ es negativa, por tanto las dos densidades son iguales casi seguro. Esto implica que $\theta^* = \theta_0$ y $\Lambda^*(\cdot) = \Lambda_0^0(\cdot)$ en $[0, \tau]$. De este modo se ha probado que $\tilde{\theta}_n \rightarrow \theta_0$ y $\hat{\Lambda}_{\tilde{\theta}_n}(x) \rightarrow \Lambda_0^0(x)$ casi seguro con $x \in [0, \tau]$. □

Nótese que como consecuencia de este teorema se cumple la condición (1.3.33) del Teorema 1.3.10. Por último, para verificar la condición (1.3.34) se demuestra el siguiente lema.

Lema 2.6.7. *Bajo la consistencia de los estimadores $\tilde{\boldsymbol{\theta}}_n$ y $\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(\cdot)$ del Teorema 2.6.6, se cumple*

$$P_0 \left[\dot{\ell}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_n, \hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}) \right] = o_{P_0}(\|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|)$$

Demostración. Teniendo en cuenta que el valor de la log-verosimilitud para el dato \mathbf{y} bajo el submodelo considerado es

$$\ell(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \log p_{\mathbf{t}, \Lambda_{\mathbf{t}}}(\mathbf{y})$$

donde la curva

$$\Lambda_{\mathbf{t}}(x) = \int_{[0, x]} (1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)) d\Lambda_0(s)$$

es el submodelo menos favorable, con $h_0(s) = \frac{M_{\boldsymbol{\theta}_0, \Lambda_0^0}^1(s)}{M_{\boldsymbol{\theta}_0, \Lambda_0^0}^0(s)}$ la dirección menos favorable. Entonces, se tiene

$$\begin{aligned} \ell(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) &= \delta [\log ((1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(x)) \lambda_0(x)) + \log K(\mathbf{t}, \mathbf{z})] \\ &\quad - K(\mathbf{t}, \mathbf{z}) \int_{[0, x]} (1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)) d\Lambda_0(s), \end{aligned}$$

y la función “score” del modelo con respecto a \mathbf{t} viene dada por

$$\begin{aligned}
 \dot{\ell}(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) &= \delta \tilde{K}(\mathbf{t}, \mathbf{z}) - \Lambda_{\mathbf{t}}(x) \tilde{K}(\mathbf{t}, \mathbf{z}) K(\mathbf{t}, \mathbf{z}) - \delta \frac{h_0(x) \lambda_0(x)}{(1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(x)) \lambda_0(x)} \\
 &\quad + K(\mathbf{t}, \mathbf{z}) \int_{[0, x]} h_0(s) d\Lambda_0(s) \\
 &= \delta \tilde{K}(\mathbf{t}, \mathbf{z}) - \tilde{K}(\mathbf{t}, \mathbf{z}) K(\mathbf{t}, \mathbf{z}) \int_{[0, x]} (1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)) d\Lambda_0(s) \\
 &\quad - \frac{\delta h_0(x)}{(1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(x))} + K(\mathbf{t}, \mathbf{z}) \int_{[0, x]} h_0(s) d\Lambda_0(s) \\
 &= \delta \tilde{K}(\mathbf{t}, \mathbf{z}) - \int_0^\tau \tilde{K}(\mathbf{t}, \mathbf{z}) K(\mathbf{t}, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} (1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)) d\Lambda_0(s) \\
 &\quad - \frac{\delta h_0(x)}{(1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(x))} + \int_0^\tau K(\mathbf{t}, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} h_0(s) d\Lambda_0(s) \quad (2.6.34) \\
 &= \int_0^\tau \tilde{K}(\mathbf{t}, \mathbf{z}) dN(s) - \int_0^\tau \frac{h_0(s)}{1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)} dN(s) \\
 &\quad - \int_0^\tau \tilde{K}(\mathbf{t}, \mathbf{z}) K(\mathbf{t}, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} (1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)) d\Lambda_0(s) \\
 &\quad + \int_0^\tau K(\mathbf{t}, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} h_0(s) d\Lambda_0(s) \\
 &= \int_0^\tau \left(\tilde{K}(\mathbf{t}, \mathbf{z}) - \frac{h_0(s)}{1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)} \right) dN(s) \\
 &\quad - \int_0^\tau \left(\tilde{K}(\mathbf{t}, \mathbf{z}) - \frac{h_0(s)}{1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)} \right) \\
 &\quad \quad K(\mathbf{t}, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} (1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)) d\Lambda_0(s) \\
 &= \int_0^\tau \left(\tilde{K}(\mathbf{t}, \mathbf{z}) - \frac{h_0(s)}{1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)} \right) dM(s),
 \end{aligned}$$

con $dN(s) = d\mathbb{1}_{\{X \leq s, \delta=1\}}$ y $dM(s) = dN(s) - K(\mathbf{t}, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} (1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)) d\Lambda_0(s)$.

De (2.6.34) se tiene

$$\begin{aligned}
 P_0[\dot{\ell}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_n, \hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n})] &= P_0 \left[\delta \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) - \frac{\delta h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} \right. \\
 &\quad - \int_0^\tau \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} (1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(s)) d\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(s) \\
 &\quad \left. + \int_0^\tau K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} h_0(s) d\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(s) \right] \\
 &= P_0 \left[- \int_0^\tau \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} d\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(s) \right. \\
 &\quad - \int_0^\tau \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(s) d\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(s) \\
 &\quad + \int_0^\tau K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} h_0(s) d\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(s) \\
 &\quad \left. + \frac{\delta \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) + \delta \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x) - \delta h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} \right] \\
 &= P_0 \left[\int_0^\tau K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} (h_0(s) - \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z})) d\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(s) \right. \\
 &\quad - \int_0^\tau \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(s) d\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(s) \\
 &\quad \left. + \frac{\delta (\tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) - h_0(x))}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} + \frac{\delta \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} \right].
 \end{aligned}$$

Nótese que

$$\begin{aligned}
 & P_0 \left[K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq x\}} (h_0(x) - \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z})) \right] \\
 &= P_0 \left[K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq x\}} h_0(x) - K(\boldsymbol{\theta}_0, \mathbf{z}) \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq x\}} \right] \\
 &= E_0 \left[K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq x\}} \right] \frac{M_{\boldsymbol{\theta}_0, \Lambda_0^0}^1(x)}{M_{\boldsymbol{\theta}_0, \Lambda_0^0}^0(x)} - E_0 \left[K(\boldsymbol{\theta}_0, \mathbf{z}) K_1(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq x\}} \right] \\
 &= M_{\boldsymbol{\theta}_0, \Lambda_0^0}^0(x) \frac{M_{\boldsymbol{\theta}_0, \Lambda_0^0}^1(x)}{M_{\boldsymbol{\theta}_0, \Lambda_0^0}^0(x)} - M_{\boldsymbol{\theta}_0, \Lambda_0^0}^1(x) \\
 &= 0.
 \end{aligned}$$

Además,

$$\begin{aligned}
 & P_0 \left[\frac{\delta(\tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) - h_0(x))}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} \right] \\
 &= \int_0^\tau \int_{\mathbb{R}^d} \frac{(\tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) - h_0(x))}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} \lambda_0^0(x) K(\boldsymbol{\theta}_0, \mathbf{z}) S_{T|\mathbf{Z}}(x|\mathbf{z}) S_{C|\mathbf{Z}}(x|\mathbf{z}) p(\mathbf{z}) d\mathbf{z} dx \\
 &= \int_0^\tau \int_{\mathbb{R}^d} \frac{(\tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) - h_0(x))}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} K(\boldsymbol{\theta}_0, \mathbf{z}) S_{T|\mathbf{Z}}(x|\mathbf{z}) S_{C|\mathbf{Z}}(x|\mathbf{z}) p(\mathbf{z}) d\mathbf{z} d\Lambda_0^0(x) \\
 &= \int_0^\tau \int_{\mathbb{R}^d} \frac{(\tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) - h_0(x))}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} K(\boldsymbol{\theta}_0, \mathbf{z}) P[T \wedge C > x|\mathbf{z}] p(\mathbf{z}) d\mathbf{z} d\Lambda_0^0(x) \\
 &= \int_0^\tau \int_{\mathbb{R}^d} \left(\frac{\tilde{K}(\boldsymbol{\theta}_0, \mathbf{z})}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} K(\boldsymbol{\theta}_0, \mathbf{z}) P[T \wedge C > x|\mathbf{z}] p(\mathbf{z}) d\mathbf{z} \right. \\
 &\quad \left. - \frac{h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} K(\boldsymbol{\theta}_0, \mathbf{z}) P[T \wedge C > x|\mathbf{z}] p(\mathbf{z}) d\mathbf{z} \right) d\Lambda_0^0(x) \\
 &= \int_0^\tau \left(\frac{M_{\boldsymbol{\theta}_0, \Lambda_0^0}^1(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} - \frac{M_{\boldsymbol{\theta}_0, \Lambda_0^0}^1(x)}{M_{\boldsymbol{\theta}_0, \Lambda_0^0}^0(x)} \frac{M_{\boldsymbol{\theta}_0, \Lambda_0^0}^0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} \right) d\Lambda_0^0(x) \\
 &= 0.
 \end{aligned}$$

También se obtiene

$$\begin{aligned}
 P_0 \left[\frac{\delta \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z})(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} \right] &= P_0 \left[\frac{\delta \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) \frac{E_0[K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq x\}}]}{E_0[K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq x\}}]} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} \right] \\
 &= P_0 \left[\frac{\delta \frac{E_0[\tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq x\}}]}{E_0[K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq x\}}]} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} \right] \\
 &= P_0 \left[\frac{\delta h_0(x)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} \right].
 \end{aligned}$$

Entonces,

$$\begin{aligned}
 P_0[\dot{\ell}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_n, \hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n})] &= P_0 \left[- \int_0^\tau \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(s) d\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(s) \right. \\
 &\quad \left. + \frac{\delta h_0(x)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} \right] \quad (2.6.35)
 \end{aligned}$$

Como

$$M(s) = N(s) - \int_0^s K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} (1 + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' h_0(s)) d\Lambda_0^0(s)$$

es una martingala con media 0, se tiene

$$\begin{aligned}
 P_0[\dot{\ell}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_n, \Lambda_0^0)] &= 0 \\
 &= P_0 \left[- \int_0^\tau \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(s) d\Lambda_0^0(s) \right. \\
 &\quad \left. + \delta \frac{h_0(x)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} \right]. \quad (2.6.36)
 \end{aligned}$$

Finalmente de las expresiones (2.6.35) y (2.6.36) se obtiene el resultado

$$\begin{aligned}
 & P_0[\dot{\ell}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_n, \hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n})] \\
 &= P_0[\dot{\ell}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_n, \hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n})] - P_0[\dot{\ell}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_n, \Lambda_0^0)] \\
 &= P_0 \left[- \int_0^\tau \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) K(\boldsymbol{\theta}_0, \mathbf{z}) \mathbb{1}_{\{X \geq s\}} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(s) d(\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(s) - \Lambda_0^0(s)) \right. \\
 &\quad \left. + \frac{2\delta h_0(x)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)} \right] \\
 &= o_{P_0}(\|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|).
 \end{aligned}$$

□

Como consecuencia del Lema anterior se cumple la condición (1.3.34) del Teorema 1.3.10. Finalmente, se puede demostrar el siguiente teorema.

Teorema 2.6.8. *Bajo las Condiciones 2.6.1, la solución $\tilde{\boldsymbol{\theta}}_n$, para una muestra de n observaciones de la función de máxima verosimilitud “profile” dada en (2.6.14) del modelo generalizado de riesgos proporcionales (2.6.13), se distribuye como*

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(\mathbf{0}, \tilde{\mathbf{I}}_0^{-1}),$$

donde $\boldsymbol{\theta}_0$ es el verdadero valor del parámetro del modelo y, además $\tilde{\mathbf{I}}_0$ viene dado por $P_0[(\dot{\ell}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0, \Lambda_0^0))(\dot{\ell}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0, \Lambda_0^0))']$.

Demostración. Como consecuencia de que se cumplen las condiciones (1.3.31)–(1.3.34) para el modelo semiparamétrico generalizado de riesgos proporcionales (2.6.13) con datos con censura por la derecha y considerando los Lemas 2.6.2 y 2.6.3, se puede aplicar el Teorema 1.3.10 y por tanto, por vi) de la Condición 2.6.1, el Corolario 1.3.11 también se cumple. Entonces se tiene que $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ se distribuye asintóticamente como $\mathcal{N}(\mathbf{0}, \tilde{\mathbf{I}}_0^{-1})$, donde $\tilde{\boldsymbol{\theta}}_n$ es el estimador de máxima log-verosimilitud “profile” dado en (2.6.14) del verdadero parámetro $\boldsymbol{\theta}_0$ del modelo (2.6.13), además $\tilde{\mathbf{I}}_0 = P_0[(\dot{\ell}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0, \Lambda_0^0))(\dot{\ell}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0, \Lambda_0^0))']$. □

De acuerdo al teorema anterior, se obtiene la distribución normal asintótica del estimador $\tilde{\boldsymbol{\theta}}_n$ de máxima log-verosimilitud “profile” del modelo (2.6.13). No obstante, en la Sección 2.6.1 se ha propuesto un estimador de máxima log-verosimilitud parcial del mismo modelo. El siguiente lema establece la equivalencia de dichos estimadores.

Lema 2.6.9. *Los estimadores de máxima verosimilitud parcial y máxima verosimilitud “profile” del modelo generalizado de riesgos proporcionales (2.6.13) coinciden.*

Demostración. Si se definen la función $N_i(x) = \mathbb{1}_{\{X_i \leq x, \delta_i=1\}}$, la función de la verosimilitud parcial se puede escribir como:

$$\mathcal{L}_n(\boldsymbol{\theta}) = \prod_{i=1}^n \prod_{0 \leq x \leq \tau} \left(\frac{\mathbb{1}_{\{X_i \geq x\}} K(\boldsymbol{\theta}, \mathbf{z}_i)}{\sum_{j=1}^n \mathbb{1}_{\{X_j \geq x\}} K(\boldsymbol{\theta}, \mathbf{z}_j)} \right)^{\Delta N_i(x)},$$

por tanto, la función de log-verosimilitud parcial viene dada por

$$\begin{aligned} \ell_n(\boldsymbol{\theta}) &= \log(\mathcal{L}_n(\boldsymbol{\theta})) \\ &= \sum_{i=1}^n \int_0^\tau \log \left(\frac{\mathbb{1}_{\{X_i \geq x\}} K(\boldsymbol{\theta}, \mathbf{z}_i)}{\sum_{j=1}^n \mathbb{1}_{\{X_j \geq x\}} K(\boldsymbol{\theta}, \mathbf{z}_j)} \right) dN_i(x) \\ &= \sum_{i=1}^n \int_0^\tau \left(\log [\mathbb{1}_{\{X_i \geq x\}} K(\boldsymbol{\theta}, \mathbf{z}_i)] - \log \left[\sum_{j=1}^n \mathbb{1}_{\{X_j \geq x\}} K(\boldsymbol{\theta}, \mathbf{z}_j) \right] \right) dN_i(x). \end{aligned}$$

Por una parte, la función “score” de la función anterior es

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}) &= \sum_{i=1}^n \int_0^\tau \left(\frac{\mathbb{1}_{\{X_i \geq x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}_i) K(\boldsymbol{\theta}, \mathbf{z}_i)}{\mathbb{1}_{\{X_i \geq x\}} K(\boldsymbol{\theta}, \mathbf{z}_i)} - \frac{\sum_{j=1}^n \mathbb{1}_{\{X_j \geq x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}_j) K(\boldsymbol{\theta}, \mathbf{z}_j)}{\sum_{j=1}^n \mathbb{1}_{\{X_j \geq x\}} K(\boldsymbol{\theta}, \mathbf{z}_j)} \right) dN_i(x) \\ &= \sum_{i=1}^n \int_0^\tau \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}_i) - \frac{n E_n [\mathbb{1}_{\{X \geq x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z})]}{n E_n [\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \right) dN_i(x) \\ &= n E_n \left[\int_0^\tau \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{E_n [\mathbb{1}_{\{X \geq x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z})]}{E_n [\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \right) dN(x) \right], \end{aligned} \tag{2.6.37}$$

donde $E_n[f]$ denota el promedio empírico de la función f , es decir $E_n[f] = \mathbb{P}_n f$.

Por otro lado, si se considera el segundo término la función “score” eficiente (2.6.24) del modelo semiparamétrico para el conjunto de n datos y considerando la función de densidad empírica se tiene

$$\begin{aligned}
 & \mathbb{P}_n \left[-K(\boldsymbol{\theta}, \mathbf{z}) \int_{[0,x]} \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{E_n[\mathbb{1}_{\{X \geq x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z})]}{E_n[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \right) d\hat{\Lambda}_0(s) \right] \\
 &= -\mathbb{P}_n \left[\int_{[0,x]} K(\boldsymbol{\theta}, \mathbf{z}) \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{E_n[\mathbb{1}_{\{X \geq x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z})]}{E_n[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \right) d\hat{\Lambda}_0(s) \right] \\
 &= -\int_0^\tau \mathbb{P}_n \left[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z}) \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{E_n[\mathbb{1}_{\{X \geq x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z})]}{E_n[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \right) \right] d\hat{\Lambda}_0(s) \\
 &= -\int_0^\tau \left(E_n \left[\mathbb{1}_{\{X \geq x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z}) \right] - \right. \\
 &\quad \left. E_n \left[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z}) \right] \frac{E_n[\mathbb{1}_{\{X \geq x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z})]}{E_n[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \right) d\hat{\Lambda}_0(s) \\
 &= 0,
 \end{aligned}$$

con $\hat{\Lambda}_0(\cdot)$ un estimador de la función de riesgo acumulativa. Además, tomando el primer término de la expresión (2.6.24) se tiene

$$\begin{aligned}
 & \mathbb{P}_n \left[\delta \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{E_n[\mathbb{1}_{\{X \geq x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z})]}{E_n[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \right) \right] \\
 &= \mathbb{P}_n \left[\int_0^\tau \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{E_n[\mathbb{1}_{\{X \geq x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z})]}{E_n[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \right) d\mathbb{1}_{\{X \leq x, \delta=1\}} \right] \quad (2.6.38) \\
 &= E_n \left[\int_0^\tau \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{E_n[\mathbb{1}_{\{X \geq x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z})]}{E_n[\mathbb{1}_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \right) dN(x) \right].
 \end{aligned}$$

Entonces la solución de las expresiones (2.6.37) y (2.6.38) es la misma. \square

Por tanto, se ha demostrado que el estimador de máxima verosimilitud parcial

(2.6.12) tiene una distribución asintótica normal. No obstante, a continuación, se estudiará el comportamiento del estimador $\hat{\theta}_n$ para muestras finitas mediante un estudio de simulación.

2.6.4. Estudio de Simulación

En este apartado, se estudian las propiedades de los estimadores de los efectos de las covariables para muestras pequeñas o moderadas mediante simulación. En concreto, será interesante valorar la normalidad de los estimadores probada en la sección anterior, pero para muestras finitas de individuos.

Se considera que el verdadero modelo de tiempo de supervivencia es el dado por (2.5.7) con parámetros $\gamma = 3.2$ y $\nu = 1.1$ para la función Weibull. La forma de generar una muestra de tiempos de supervivencia de este modelo es la explicada en la Sección 2.4.

Se consideran tamaños de muestra $n = 100, 200$ y 400 y diferentes proporciones de censura como 15 %, 45 % y 70 %, para un total de 500 bases de datos simuladas en cada configuración. Se han hecho experimentos para diferentes tipos de covariables. Los resultados se presentan para los siguientes diseños, bajo los cuales se generan los tiempos de supervivencia.

Diseño 1: Se considera $\beta'_0 \mathbf{Z} = \sum_{i=1}^4 \beta_{0i} Z_i$ con $\beta_0 = (0.5, 0.2, 0.3, 0.1)$, $a = 8$ y Z_1, Z_2, Z_3 covariables binarias generadas de un factor de 4 niveles con probabilidad de $\{0.37, 0.19, 0.30, 0.14\}$ para cada nivel y $Z_4 \sim \mathcal{U}(-1, 1)$.

Diseño 2: Se toma $\beta'_0 \mathbf{Z} = \sum_{i=1}^3 \beta_{0i} Z_i$ con $\beta_0 = (0.3, 0.6, 0.1)$, $a = 8$, las covariables $Z_1 \sim \mathcal{B}(0.35)$, $Z_2 \sim \mathcal{U}(-1, 1)$ y $Z_3 \sim \mathcal{N}(1.5, 2)$.

Una vez generada cada base de datos, se ajusta el modelo semiparamétrico generalizado de riesgos proporcionales, es decir se estima el valor del vector de parámetros β y el valor del parámetro a . El parámetro a se estima bajo el procedimiento de maximización por contornos propuesto en la Subsección 2.5.3. En este

caso, se ajusta el modelo para $a \in \{7.50, 7.75, 8.00, 8.25, 8.50\}$ y se selecciona el modelo con valor de a tal que tenga mayor log-verosimilitud. En las tablas siguientes, se muestra la media y la mediana de la estimación de los parámetros β 's, así como la media del error estándar (EE) estimado y el EE empírico, el p -valor para el test de normalidad de Shapiro-Wilks y por último, se muestra la probabilidad empírica de cubrimiento (PC) para valores nominales del 90 %, 95 % y 99 %.

En las Tablas 2.7 y 2.8 se muestran los resultados de los experimentos para el Diseño 1 y 2, respectivamente. En este caso, se observa que la media y la mediana del valor de la estimación se aproxima al verdadero valor del parámetro cuando el tamaño de la muestra n crece, y que el EE empírico y estimado son cercanos y disminuyen cuando n crece. Además, el p -valor del test de normalidad es mayor para n grande, lo cual corrobora el resultado asintótico de la sección anterior sobre la normalidad de los estimadores de los efectos de las covariables. Como es de esperar, la probabilidad empírica de cobertura es más próxima al valor nominal cuando n crece. También se observa que cuando se tiene un porcentaje pequeño de datos censurados los resultados son mejores y éstos empeoran cuando dicho porcentaje crece. Las conclusiones obtenidas son las mismas para ambos diseños.

Finalmente, en las Figuras 2.3, 2.4 y 2.5, se muestran las gráficas de la función de densidad empírica de los parámetros estimados bajo el Diseño 1 y 2, para diferentes valores de censura por la derecha. Se observa que las gráficas de la densidad empírica se acercan a la gráfica de una distribución normal con media en el verdadero valor del parámetro cuando n crece. Para los diferentes porcentajes de censura la forma de la gráfica se mantiene, sin embargo, la altura disminuye, es decir, tienen varianza mayor al aumentar el porcentaje de censura.

Modelo Generalizado para Datos con Censura por la Derecha

Cens.	n	Par.	Media	Mediana	EE	EEe	p-valor	PC para .90	PC para .95	PC para .99
15 %	100	β_1	0.52896	0.52070	0.12460	0.13066	0.00002	0.872	0.924	0.976
		β_2	0.21965	0.21197	0.08306	0.08491	0.00005	0.872	0.930	0.980
		β_3	0.32589	0.32140	0.11383	0.11941	0.00000	0.914	0.952	0.986
		β_4	0.10197	0.09954	0.05665	0.05979	0.01339	0.884	0.942	0.984
	200	β_1	0.51267	0.49862	0.08451	0.09725	0.00000	0.846	0.912	0.978
		β_2	0.20341	0.20339	0.05660	0.05903	0.17780	0.878	0.942	0.984
		β_3	0.31225	0.30518	0.07773	0.08479	0.00070	0.890	0.938	0.978
		β_4	0.09996	0.10109	0.03870	0.03866	0.19144	0.904	0.948	0.986
	400	β_1	0.50603	0.50043	0.05899	0.06881	0.00854	0.822	0.898	0.978
		β_2	0.20462	0.20295	0.03981	0.04011	0.04824	0.904	0.954	0.986
		β_3	0.30716	0.30274	0.05424	0.05973	0.00303	0.862	0.922	0.980
		β_4	0.10179	0.10170	0.02713	0.02740	0.78848	0.876	0.934	0.990
45 %	100	β_1	0.54146	0.52331	0.16010	0.17754	0.00000	0.858	0.900	0.956
		β_2	0.21324	0.20800	0.10441	0.11171	0.00018	0.870	0.932	0.982
		β_3	0.32191	0.30361	0.14705	0.15981	0.00000	0.878	0.916	0.974
		β_4	0.11144	0.10794	0.07221	0.07659	0.16895	0.866	0.934	0.976
	200	β_1	0.52802	0.51629	0.10798	0.11714	0.00066	0.856	0.918	0.972
		β_2	0.20995	0.20793	0.07161	0.07443	0.39918	0.886	0.938	0.986
		β_3	0.31986	0.31547	0.09907	0.10533	0.02710	0.862	0.928	0.978
		β_4	0.10801	0.10607	0.04909	0.04939	0.33180	0.902	0.946	0.984
	400	β_1	0.50671	0.50083	0.07331	0.08141	0.00150	0.860	0.912	0.978
		β_2	0.20492	0.20530	0.04959	0.05388	0.01724	0.896	0.938	0.982
		β_3	0.30805	0.30543	0.06759	0.07702	0.00389	0.868	0.918	0.978
		β_4	0.10229	0.10410	0.03383	0.03402	0.79969	0.900	0.942	0.984
70 %	100	β_1	0.56340	0.52049	0.44202	0.51533	0.00000	0.834	0.884	0.938
		β_2	0.24258	0.20871	0.28117	0.47029	0.00000	0.864	0.908	0.962
		β_3	0.30960	0.29561	0.63577	0.62722	0.00000	0.860	0.914	0.956
		β_4	0.10420	0.09984	0.10412	0.16359	0.00000	0.866	0.912	0.964
	200	β_1	0.52015	0.50629	0.14801	0.15207	0.00002	0.866	0.926	0.984
		β_2	0.21102	0.20937	0.09808	0.10264	0.04140	0.894	0.940	0.974
		β_3	0.30352	0.31030	0.19855	0.19320	0.00000	0.900	0.940	0.972
		β_4	0.10836	0.10876	0.06750	0.06929	0.48689	0.874	0.926	0.982
	400	β_1	0.51422	0.50877	0.10090	0.11539	0.02883	0.842	0.902	0.962
		β_2	0.20797	0.20566	0.06760	0.07152	0.43157	0.872	0.922	0.982
		β_3	0.30877	0.31168	0.09271	0.09495	0.59518	0.886	0.942	0.980
		β_4	0.10019	0.09903	0.04614	0.04748	0.49890	0.874	0.946	0.992

Tabla 2.7: Resultados para el Diseño 1 tomando 500 réplicas para diferente censura por la derecha y tamaño de muestra.

Modelo Generalizado para Datos con Censura por la Derecha

Cens.	n	Par.	Media	Mediana	EE	EEe	p-valor	PC para .90	PC para .95	PC para .99
15 %	100	β_1	0.31225	0.30630	0.08559	0.09181	0.00021	0.864	0.938	0.972
		β_2	0.62663	0.61607	0.09405	0.10834	0.00000	0.840	0.894	0.960
		β_3	0.10628	0.10476	0.02225	0.02382	0.00003	0.864	0.914	0.970
	200	β_1	0.30951	0.30939	0.05781	0.06059	0.00454	0.884	0.940	0.978
		β_2	0.60836	0.60294	0.06303	0.07822	0.00003	0.802	0.912	0.974
		β_3	0.10175	0.10006	0.01491	0.01739	0.00001	0.854	0.910	0.958
	400	β_1	0.30166	0.29935	0.03979	0.04496	0.00013	0.874	0.928	0.978
		β_2	0.60265	0.59819	0.04342	0.05902	0.00092	0.766	0.846	0.940
		β_3	0.10010	0.09977	0.01017	0.01121	0.17298	0.870	0.934	0.980
45 %	100	β_1	0.31524	0.31323	0.10690	0.11154	0.00026	0.882	0.950	0.982
		β_2	0.63217	0.60955	0.11788	0.12870	0.00000	0.868	0.932	0.964
		β_3	0.10680	0.10288	0.02793	0.03007	0.00000	0.852	0.922	0.972
	200	β_1	0.30726	0.29930	0.07316	0.07522	0.00038	0.872	0.932	0.982
		β_2	0.62303	0.62235	0.08059	0.09249	0.01260	0.826	0.896	0.966
		β_3	0.10453	0.10253	0.01888	0.02072	0.00094	0.844	0.898	0.972
	400	β_1	0.30600	0.30154	0.05034	0.05477	0.08605	0.866	0.926	0.984
		β_2	0.61245	0.60473	0.05492	0.06542	0.00199	0.820	0.894	0.964
		β_3	0.10237	0.10105	0.01290	0.01421	0.00007	0.862	0.908	0.966
70 %	100	β_1	0.34020	0.31814	0.16028	0.18905	0.00000	0.864	0.918	0.956
		β_2	0.67274	0.64338	0.17912	0.22422	0.00000	0.848	0.896	0.952
		β_3	0.11270	0.10424	0.04217	0.05451	0.00000	0.858	0.908	0.948
	200	β_1	0.31892	0.31371	0.10125	0.10926	0.00001	0.864	0.928	0.972
		β_2	0.62598	0.60825	0.11109	0.12647	0.00000	0.846	0.908	0.962
		β_3	0.10519	0.10257	0.02609	0.02842	0.00000	0.868	0.918	0.972
	400	β_1	0.30947	0.30308	0.06885	0.07721	0.00005	0.884	0.928	0.976
		β_2	0.62163	0.61481	0.07544	0.08963	0.00398	0.818	0.896	0.960
		β_3	0.10292	0.10166	0.01757	0.01942	0.00317	0.858	0.916	0.972

Tabla 2.8: Resultados para el Diseño 2 tomando 500 réplicas para diferente censura por la derecha y tamaño de muestra.

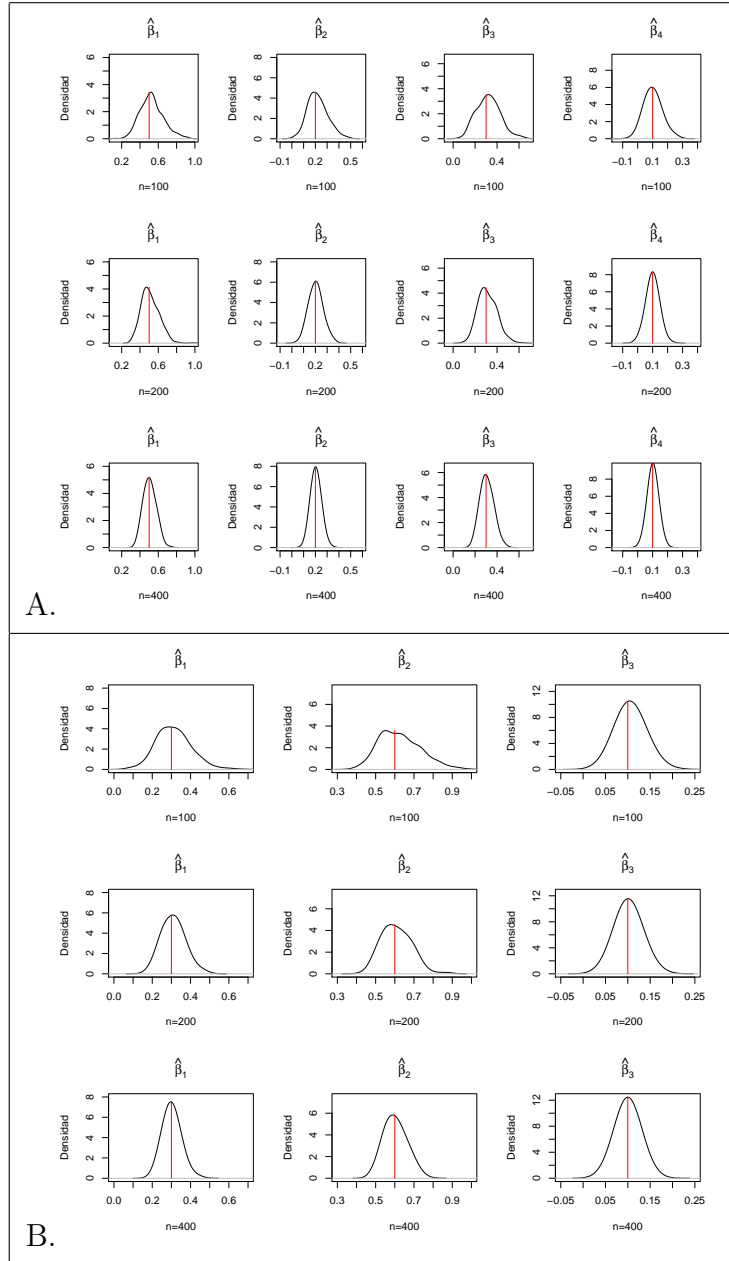


Figura 2.3: Gráfica de la densidad empírica de los parámetros estimados para 500 réplicas con 15% de censura por la derecha para diferentes tamaños de muestra para el A. Diseño 1 y B. Diseño 2.

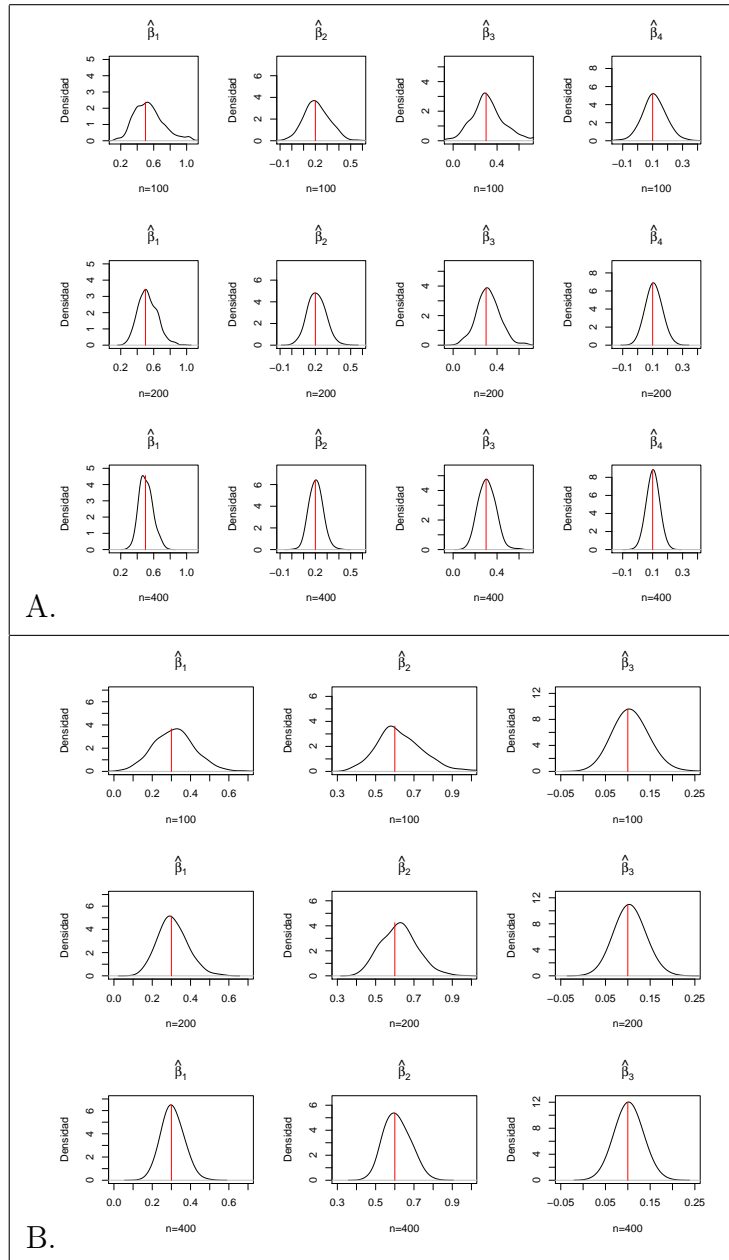


Figura 2.4: Gráfica de la densidad empírica de los parámetros estimados para 500 réplicas con 45% de censura por la derecha para diferentes tamaños de muestra para el A. Diseño 1 y B. Diseño 2.

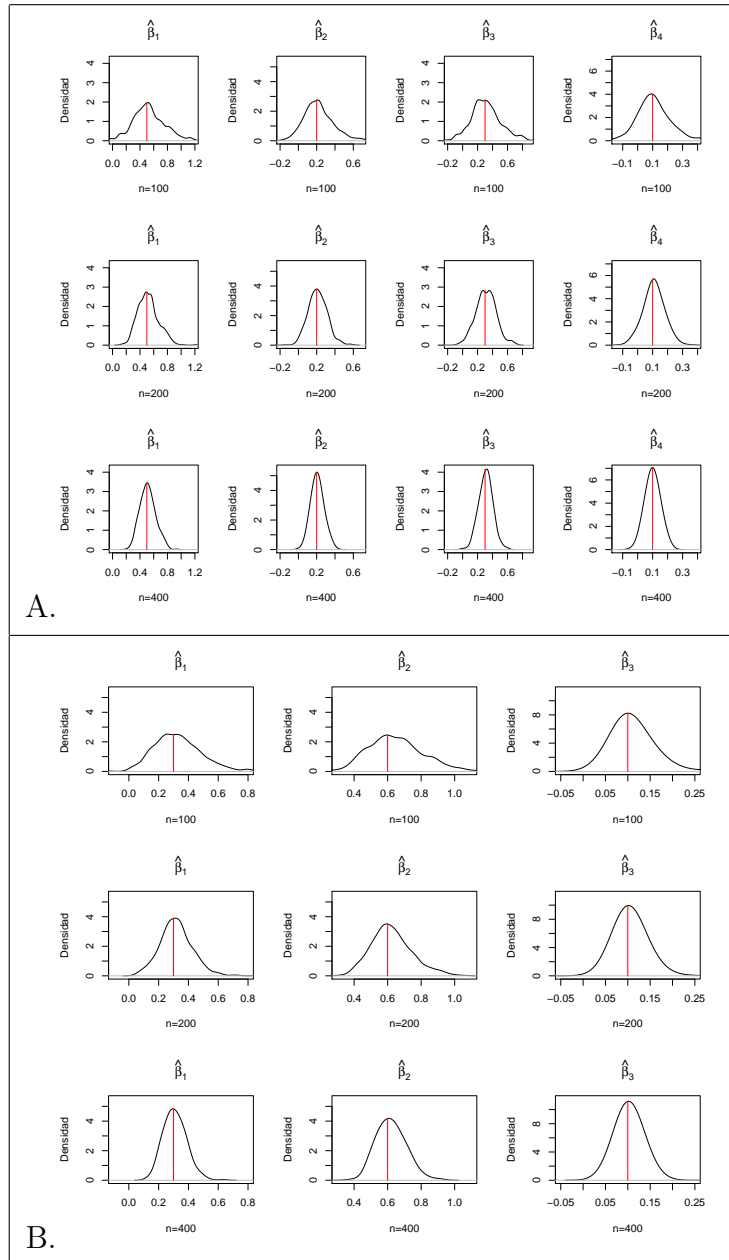


Figura 2.5: Gráfica de la densidad empírica de los parámetros estimados para 500 réplicas con 70% de censura por la derecha para diferentes tamaños de muestra para el A. Diseño 1 y B. Diseño 2.

2.6.5. Estimadores de la Función de Supervivencia

En esta sección se obtendrá un estimador de la función de riesgo base $\Lambda_0(\cdot)$ para el modelo generalizado de riesgos proporcionales (2.6.13). Una vez que se tenga el estimador $\hat{\Lambda}_0(\cdot)$, será usado para proponer estimadores de la función de supervivencia para el modelo antes mencionado.

Se sabe que la función de verosimilitud para una muestra con n datos con posible censura por la derecha es

$$\begin{aligned} L_n(\boldsymbol{\theta}, \Lambda_0) &\propto \prod_{i=1}^n [\lambda_0(x_i) K(\boldsymbol{\theta}, \mathbf{z}_i)]^{\delta_i} \exp[-\Lambda_0(x_i) K(\boldsymbol{\theta}, \mathbf{z}_i)] \\ &= \left(\prod_{j=1}^D [\lambda_0(x_j) K(\boldsymbol{\theta}, \mathbf{z}_j)] \right) \left(\prod_{i=1}^n \exp[-\Lambda_0(x_i) K(\boldsymbol{\theta}, \mathbf{z}_i)] \right) \\ &= \left(\prod_{j=1}^D [\lambda_0(x_j) K(\boldsymbol{\theta}, \mathbf{z}_j)] \right) \left(\exp \left[- \sum_{i=1}^n \Lambda_0(x_i) K(\boldsymbol{\theta}, \mathbf{z}_i) \right] \right) \end{aligned}$$

donde x_j con $j = 1, \dots, D$ corresponden a los tiempos sin censura. Supóngase que el valor de $\boldsymbol{\theta}$ es fijo y se define

$$\Lambda_0(x) = \sum_{x^* < x} \lambda_0(x^*) \quad (2.6.39)$$

entonces se tiene

$$L_n(\lambda_0) \propto \left(\prod_{j=1}^D \lambda_0(x_j) \right) \left(\exp \left[- \sum_{i=1}^n K(\boldsymbol{\theta}, \mathbf{z}_i) \left(\sum_{x^* < x_i} \lambda_0(x^*) \right) \right] \right)$$

considerando $\lambda_0(x) = 0$ cuando se tiene un dato censurado se tiene

$$\begin{aligned} L_n(\lambda_0) &\propto \left(\prod_{j=1}^D \lambda_0(x_j) \right) \left(\exp \left(- \sum_{j=1}^D \lambda_0(x_j) \sum_{i=1}^n \mathbb{1}_{\{X_i \geq x_j\}} K(\boldsymbol{\theta}, \mathbf{z}_i) \right) \right) \\ &= \left(\prod_{j=1}^D \lambda_0(x_j) \right) \left(\prod_{j=1}^D \exp \left(- \lambda_0(x_j) \sum_{i=1}^n \mathbb{1}_{\{X_i \geq x_j\}} K(\boldsymbol{\theta}, \mathbf{z}_i) \right) \right) \\ &= \prod_{j=1}^D \lambda_0(x_j) \exp \left(- \lambda_0(x_j) \sum_{i=1}^n \mathbb{1}_{\{X_i \geq x_j\}} K(\boldsymbol{\theta}, \mathbf{z}_i) \right). \end{aligned}$$

Para obtener un estimador de $\lambda_0(\cdot)$, suponiendo que el máximo existe, se deriva $L_n(\lambda_0)$ con respecto a cada $\lambda_0(x_j)$ y se obtiene

$$\begin{aligned} \frac{\partial L_n(\lambda_0)}{\partial \lambda_0(x_j)} = & \exp \left(-\lambda_0(x_j) \sum_{i=1}^n \mathbb{1}_{\{X_i \geq x_j\}} K(\boldsymbol{\theta}, \mathbf{z}_i) \right) \\ & + \lambda_0(x_j) \exp \left(-\lambda_0(x_j) \sum_{i=1}^n \mathbb{1}_{\{X_i \geq x_j\}} K(\boldsymbol{\theta}, \mathbf{z}_i) \right) \left(-\sum_{i=1}^n \mathbb{1}_{\{X_i \geq x_j\}} K(\boldsymbol{\theta}, \mathbf{z}_i) \right) \end{aligned}$$

igualando a cero se obtiene el máximo, si existe, en

$$\lambda_0(x_j) = \frac{1}{\sum_{i=1}^n \mathbb{1}_{\{X_i \geq x_j\}} K(\boldsymbol{\theta}, \mathbf{z}_i)}$$

por tanto, por la expresión (2.6.39) se tiene un estimador para $\Lambda_0(\cdot)$ y viene dado por

$$\hat{\Lambda}_0(t) = \sum_{t^* < t} \frac{1}{w(t^*)} \quad (2.6.40)$$

donde

$$w(t^*) = \sum_{i=1}^n \mathbb{1}_{\{X_i \geq t^*\}} K(\boldsymbol{\theta}, \mathbf{z}_i). \quad (2.6.41)$$

El estimador dado en la ecuación (2.6.40) es el equivalente al estimador de Breslow de la función de riesgo base $\Lambda_0(\cdot)$ en el modelo de Cox. En base a este estimador se pueden proponer los siguientes estimadores para la función de supervivencia para un valor de covariables \mathbf{z}^* dado y los valores estimados $\hat{\boldsymbol{\theta}}$ para los parámetros.

Primera propuesta: Usando la ecuación (2.2.5) y sustituyendo el estimador (2.6.40) se tiene

$$\begin{aligned} \hat{S}_1(t|\hat{\boldsymbol{\theta}}, \mathbf{z}^*) &= \exp\{-K(\hat{\boldsymbol{\theta}}, \mathbf{z}^*)\hat{\Lambda}_0(t)\} \\ &= \exp\{-\hat{\Lambda}_0(t)\}^{K(\hat{\boldsymbol{\theta}}, \mathbf{z}^*)} \\ &= \exp\left\{-\sum_{t^* < t} \frac{1}{w(t^*)}\right\}^{K(\hat{\boldsymbol{\theta}}, \mathbf{z}^*)} \end{aligned} \quad (2.6.42)$$

con $w(\cdot)$ como en (2.6.41).

Segunda propuesta: De la expresión (2.6.40) se obtiene

$$\exp[\hat{\Lambda}_0(t)] = \exp \left[\sum_{t^* < t} \frac{1}{w(t^*)} \right],$$

usando la definición de producto integral se tiene

$$\exp[\hat{\Lambda}_0(t)] = \prod_{t^* < t} \left(1 + \frac{1}{w(t^*)} \right) \quad (2.6.43)$$

y por otro lado, de la expresión (2.2.5) se puede obtener

$$S(t|\boldsymbol{\theta}, \mathbf{z}) = [\exp[\Lambda_0(t)]]^{-K(\boldsymbol{\theta}, \mathbf{z})}.$$

Así, sustituyendo el valor dado (2.6.43) en la expresión anterior se tiene el siguiente estimador

$$\hat{S}_2(t|\hat{\boldsymbol{\theta}}, \mathbf{z}^*) = \left[\prod_{t^* < t} \left(1 + \frac{1}{w(t^*)} \right) \right]^{-K(\hat{\boldsymbol{\theta}}, \mathbf{z}^*)} \quad (2.6.44)$$

con $w(\cdot)$ como en (2.6.41).

Tercera propuesta: Multiplicando la expresión (2.6.40) por $-K(\hat{\boldsymbol{\theta}}, \mathbf{z}^*)$ se tiene

$$\begin{aligned} -K(\hat{\boldsymbol{\theta}}, \mathbf{z}^*)\hat{\Lambda}_0(t) &= -K(\hat{\boldsymbol{\theta}}, \mathbf{z}^*) \sum_{t^* < t} \frac{1}{w(t^*)} \\ &= - \sum_{t^* < t} \frac{K(\hat{\boldsymbol{\theta}}, \mathbf{z}^*)}{w(t^*)} \end{aligned}$$

aplicando la función exponencial

$$\begin{aligned} \exp \left[-K(\hat{\boldsymbol{\theta}}, \mathbf{z}^*)\hat{\Lambda}_0(t) \right] &= \exp \left[- \sum_{t^* < t} \frac{K(\hat{\boldsymbol{\theta}}, \mathbf{z}^*)}{w(t^*)} \right] \\ &= \exp \left[\sum_{t^* < t} \frac{K(\hat{\boldsymbol{\theta}}, \mathbf{z}^*)}{w(t^*)} \right]^{-1}. \end{aligned}$$

Entonces, por la definición de producto integral se obtiene el siguiente estimador

$$\hat{S}_3(t|\hat{\boldsymbol{\theta}}, \mathbf{z}^*) = \left[\prod_{t^* < t} \left(1 + \frac{K(\hat{\boldsymbol{\theta}}, \mathbf{z}^*)}{w(t^*)} \right) \right]^{-1}. \quad (2.6.45)$$

Nótese que las funciones (2.6.42), (2.6.44) y (2.6.45) son siempre positivas, además como consecuencia del Teorema 5 de Breslow (1974), bajo las condiciones necesarias, los estimadores de la función de supervivencia (2.6.42)–(2.6.45) son asintóticamente normales con media 0.

Para estimar la varianza de $\hat{S}_i(t|\hat{\boldsymbol{\theta}}, \mathbf{z}^*)$ con $i = 1, 2, 3$ se usará el desarrollo de Taylor de primer orden como sigue

$$\hat{S}_i(t|\boldsymbol{\theta}, \mathbf{z}^*) \approx \hat{S}_i(t|\boldsymbol{\theta}, \mathbf{z}^*) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \frac{\partial}{\partial \boldsymbol{\theta}} \hat{S}_i(t|\boldsymbol{\theta}, \mathbf{z}^*) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

Obsérvese que en el primer término la variabilidad es causada por el estimador $\hat{\Lambda}_0(\cdot)$ y en el segundo término es causada por el estimador $\hat{\boldsymbol{\theta}}$. Así, como $\frac{\partial}{\partial \boldsymbol{\theta}} \hat{S}_i(t|\boldsymbol{\theta}, \mathbf{z}^*) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ es asintóticamente constante y $(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'$ es asintóticamente independiente de $\hat{S}_i(t|\boldsymbol{\theta}, \mathbf{z}^*) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, por lo tanto la varianza de $\hat{S}_i(t|\boldsymbol{\theta}, \mathbf{z}^*)$ se puede aproximar por

$$\text{var} \left[\hat{S}_i(t|\boldsymbol{\theta}, \mathbf{z}^*) \right] \approx \text{var} \left[\hat{S}_i(t|\boldsymbol{\theta}, \mathbf{z}^*) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right] + \left[\frac{\partial}{\partial \boldsymbol{\theta}} \hat{S}_i(t|\boldsymbol{\theta}, \mathbf{z}^*) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right]^2 \text{var} \left[\hat{\boldsymbol{\theta}} \right].$$

A continuación, se comparan los estimadores para la función de supervivencia antes descritos. Para dicha comparación, primero, se hará un estudio de simulación y posteriormente, se comparan usando la base de datos reales sobre cáncer de pulmón.

Comparación de los Estimadores con Datos Simulados

En esta sección, se compara el comportamiento de los estimadores de la función de supervivencia descritos en la sección anterior.

La comparación se hará mediante datos simulados, se usarán los diseños 1 y 2 descritos y usados en la Sección 2.6.4 para estudiar las propiedades de los estimadores de los parámetros.

Una vez que se han generado las bases de datos y se han estimado los parámetros $\hat{\boldsymbol{\beta}}$ y \hat{a} , fijamos el valor \mathbf{z}^* para el cual se estimará la función de supervivencia,

$\hat{S}_j(\cdot|\hat{\beta}, \hat{a}, \mathbf{z}^*)$ para $j = 1, 2, 3$. Se calcula la función de supervivencia \hat{S}_j en el tiempo t_p del p -percentil con $p = 0.9, 0.5, 0.1$.

Para comparar los estimadores $\hat{S}_j(\cdot)$ en los puntos t_p usaremos el valor del sesgo relativo definido por

$$b = 100 \frac{|\hat{E}[\hat{S}_j(t_p|\hat{\beta}, \hat{a}, \mathbf{z}^*)] - p|}{p}$$

para $j = 1, 2, 3$ y $p = 0.9, 0.5, 0.1$, donde $\hat{E}[\hat{S}_j(t_p|\hat{\beta}, \hat{a}, \mathbf{z}^*)]$ es la media del valor el estimador $\hat{S}_j(t_p|\hat{\beta}, \hat{a}, \mathbf{z}^*)$ en cada muestra l con $l = 1, \dots, 500$.

En las Figuras 2.6 y 2.7 se presenta el sesgo relativo para los experimentos con datos simulados. Se consideran experimentos con 15 %, 45 % y 70 % de censura por la derecha. Nótese que el sesgo relativo decrece cuando p incrementa, al igual que cuando incrementa el tamaño de muestra n . Para $p > 0.1$ todos los estimadores son comparables en relación al sesgo relativo de la función de supervivencia estimada, aunque en algunos casos $\hat{S}_2(\cdot)$ tiene menor sesgo que el resto. Para $p = 0.1$, se puede observar que el estimador $\hat{S}_2(\cdot)$ es claramente el mejor cuando hay suficientes datos completos, sin embargo es el peor cuando se tienen muchos datos censurados. Por tanto, este estimador es recomendable para un porcentaje de censura pequeño o moderado para el Diseño 2 y porcentaje bajo de censura para el Diseño 1.

Andersen y otros (1996) comparan cuatro estimadores de la función de supervivencia para el modelo tradicional de Cox. Al final de una comparación y discusión sobre dichos estimadores, concluyen que el estimador de la función de supervivencia que se basa en el producto integral del estimador de Breslow es la mejor alternativa, con la objeción de que este estimador puede dar valores negativos de la función de supervivencia. En el estudio presentado en este trabajo, también concluimos que el estimador de la función de supervivencia que se basa en el producto integral del equivalente al estimador de Breslow es la mejor alternativa, siempre y cuando se tengan suficientes datos completos.

A continuación, se verá el comportamiento de los estimadores con datos reales.

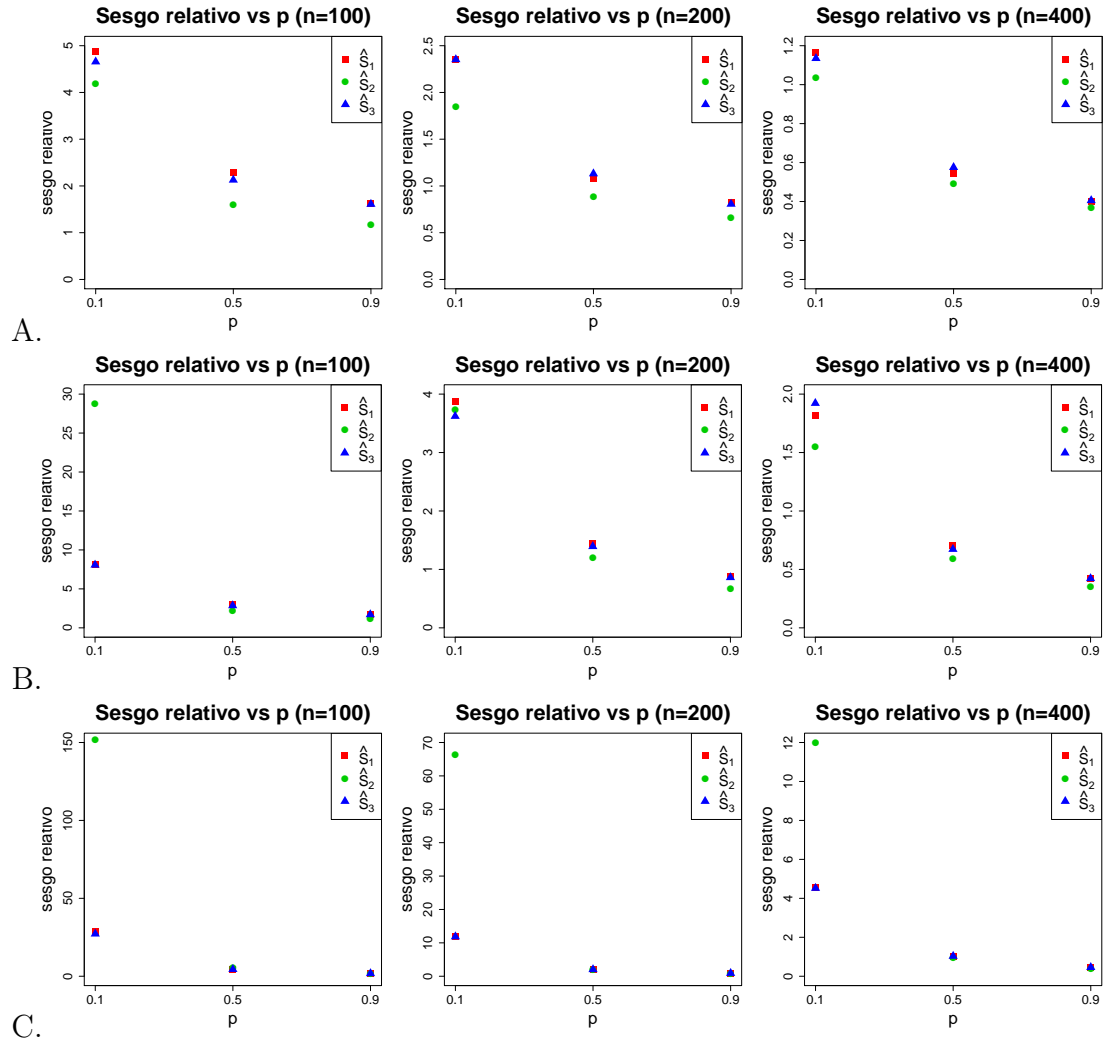


Figura 2.6: Gráfica del sesgo relativo de la función de supervivencia estimada el Diseño 1 para 500 réplicas y diferentes tamaños de muestra con A. 15 % de censura por la derecha, B. 45 % de censura por la derecha y C. 70 % de censura por la derecha.

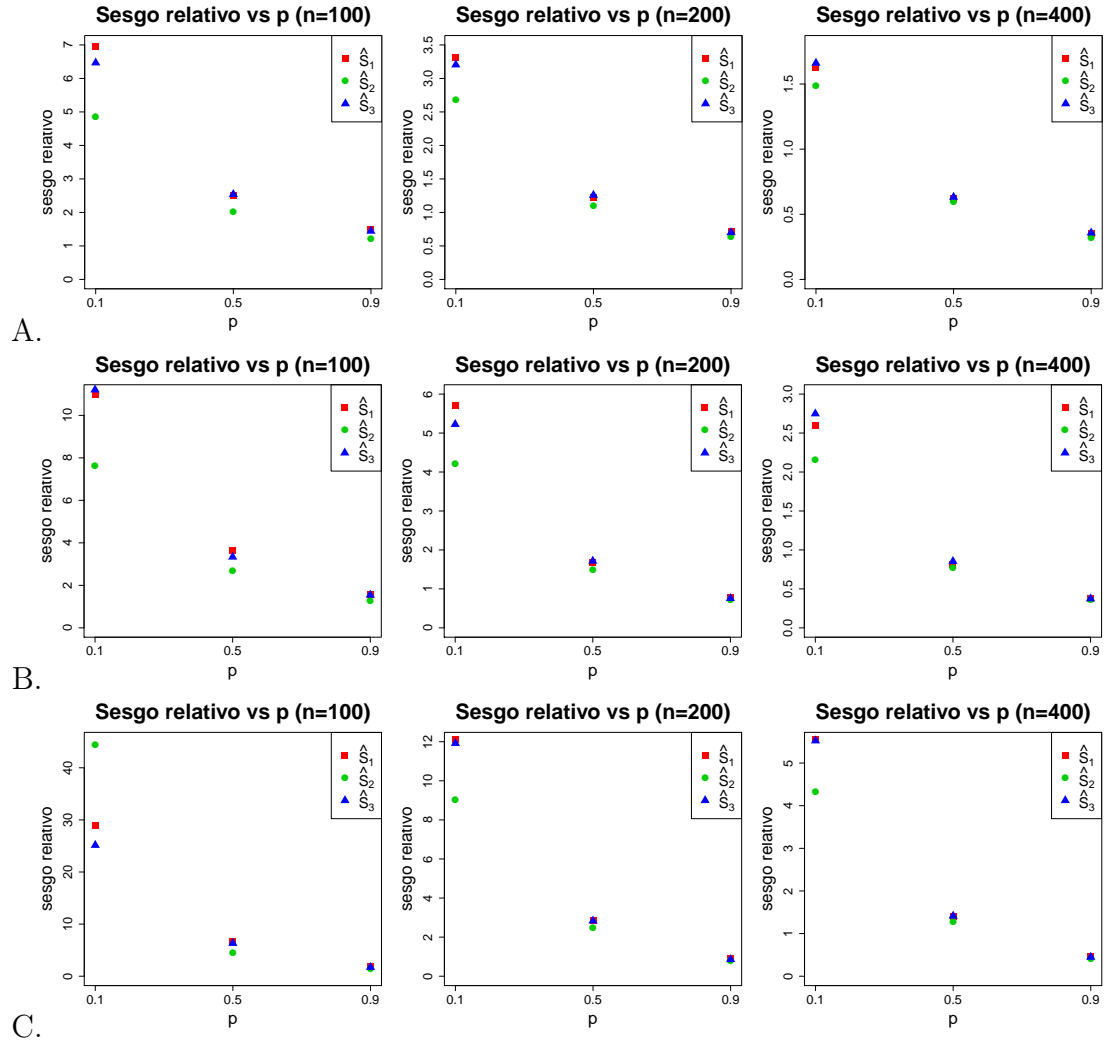


Figura 2.7: Gráfica del sesgo relativo de la función de supervivencia estimada el Diseño 2 para 500 réplicas y diferentes tamaños de muestra con A. 15 % de censura por la derecha, B. 45 % de censura por la derecha y C. 70 % de censura por la derecha.

Comparación de los Estimadores con Datos Reales

En este apartado se usan los datos sobre cáncer de pulmón estudiados por MacKenzie (1996) y en la Sección 2.6.2 de este trabajo. Estos datos se usan para comparar los estimadores de la función de supervivencia $\hat{S}_j(\cdot|\hat{\beta}, \hat{a}, \mathbf{z}^*)$, con $j = 1, 2, 3$ propuestos en la Sección 2.6.5, que a su vez se comparan con el estimador de Kaplan–Meier.

Una vez que se han estimado los parámetros $\hat{\beta} = (0.014, 0.024)$ y el parámetro $\hat{a} = 4.5$ usando maximización por contornos, como se describe en la Sección 2.6.2, se dibujan las gráficas de las estimaciones $\hat{S}_j(\cdot|\hat{\beta}, \hat{a}, \mathbf{z}^*)$, con $j = 1, 2, 3$, donde $\mathbf{z}^* = (69, 1)$ es el valor medio de las covariables \mathbf{z} de los datos de cáncer de pulmón. En la Figura 2.8 se muestran las gráficas de las funciones de dichos estimadores. Como se puede observar es difícil ver la diferencia entre los diferentes estimadores, por ello, en la Figura 2.9 se muestran 2 trozos de las mismas gráficas con la finalidad de hacer un acercamiento. En esta figura se puede ver que el estimador $\hat{S}_2(\cdot|\hat{\beta}, \hat{a}, \mathbf{z}^*)$ está más separado de los estimadores $\hat{S}_1(\cdot|\hat{\beta}, \hat{a}, \mathbf{z}^*)$ y $\hat{S}_3(\cdot|\hat{\beta}, \hat{a}, \mathbf{z}^*)$, pero más cercano al estimador de Kaplan–Meier, lo cual lo hace una mejor propuesta. Cabe mencionar que el tamaño de muestra de los datos es “grande”, y además, se tienen muy pocos datos censurados. Esta conclusión refuerza la conclusión obtenida en el estudio de simulación.

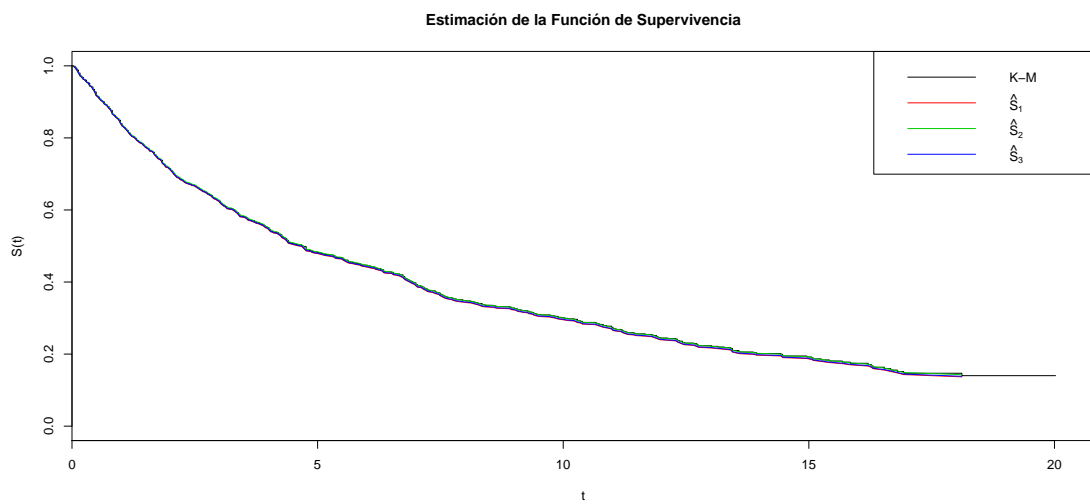


Figura 2.8: Comparación de los estimadores de la función de supervivencia \hat{S}_1 , \hat{S}_2 y \hat{S}_3 y el estimador de Kaplan–Meier usando los datos de cáncer de pulmón.

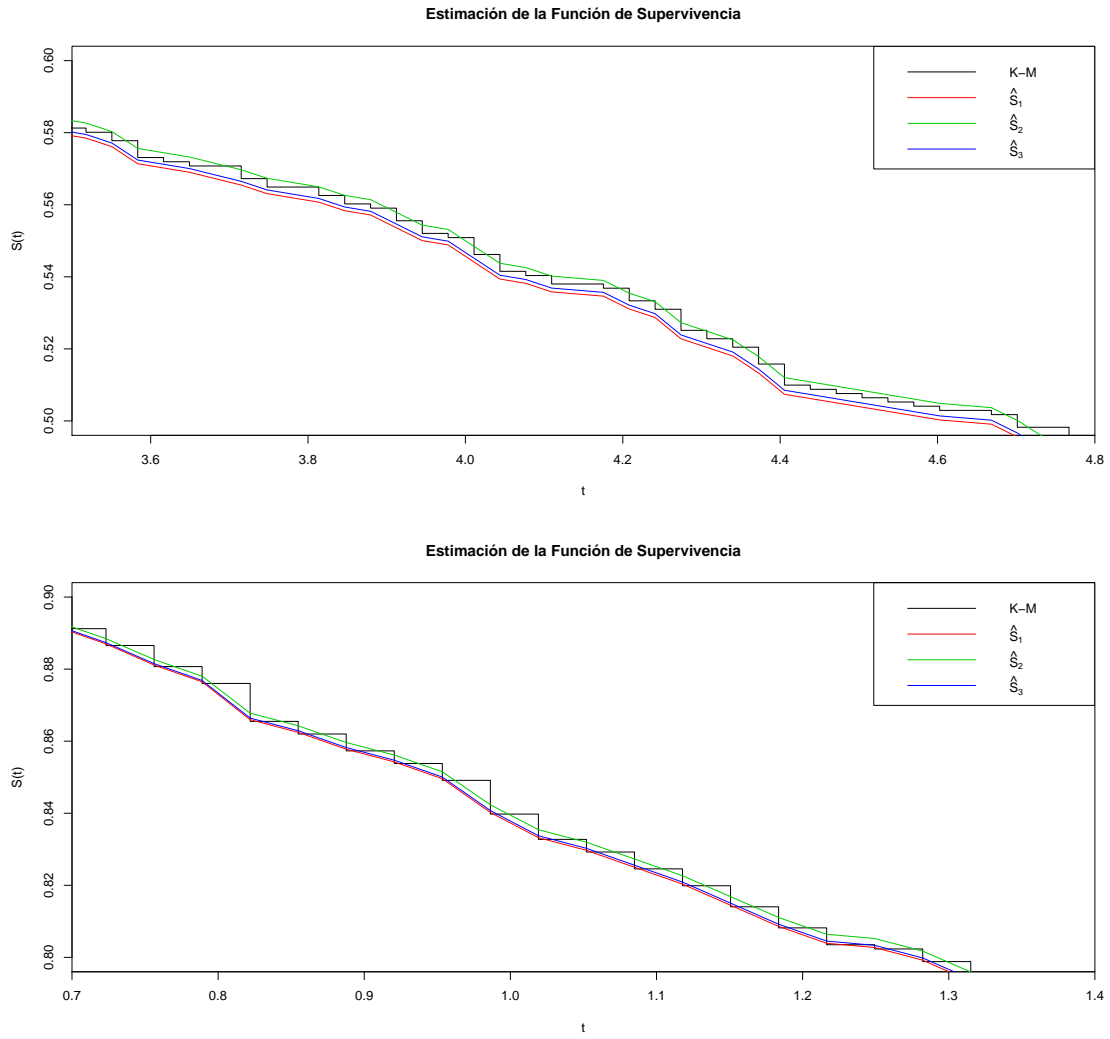


Figura 2.9: Comparación de los estimadores de la función de supervivencia \hat{S}_1 , \hat{S}_2 y \hat{S}_3 y el estimador de Kaplan–Meier usando los datos de cáncer de pulmón.

Capítulo 3

Modelo de Cox para Datos Doblemente Censurados

3.1. Introducción

En este capítulo, se estudia el ajuste del modelo de regresión de Cox dado en (1.2.12) a datos de supervivencia que pueden presentar censura por la derecha o censura por la izquierda, es decir, datos doblemente censurados. Este tipo de censura fue estudiada por primera vez por Turnbull (1974) y posteriormente por muchos otros como Chang y Yang (1987), Samuelsen (1989), Chang (1990), Huang (1999) y Chen y Zhou (2003). Véase también Gu y Zhang (1993), van der Laan y Gill (1999) y Kim (2003). Los datos doblemente censurados no deben confundirse con los datos con tiempo de supervivencia doblemente censurado estudiados por Gómez y Lagakos (1994), entre otros. Los datos doblemente censurados han sido estudiados recientemente por Patilea y Rolin (2006) y Julià y Gómez (2011), sin embargo, estos autores abordan el problema de estimar la función de supervivencia para este tipo de datos sin considerar covariables en el estudio.

Un ejemplo en el que se tienen datos doblemente censurados es un estudio

donde la variable aleatoria T , es la edad de contagio de una enfermedad no fatal con síntomas iniciales no evidentes. Los individuos son analizados sólo una vez y pueden ser categorizados como sigue:

- i) Se presenta evidencia de la enfermedad y la edad de contagio es conocida.
- ii) La enfermedad es diagnosticada pero la edad de contagio es desconocida o la información acerca de ésta es cuestionable.
- iii) La enfermedad no es diagnosticada en el momento de ser examinado.

Sea C la edad del individuo en el momento de ser examinado y A una variable aleatoria discreta con valores en $\{0, 1, 2\}$ que indican si el dato es no censurado, censurado por la derecha o censurado por la izquierda, respectivamente. En el caso i) la edad de contagio de la enfermedad T es observado, en tal caso $X = T$ y $A = 0$. En el caso ii) la edad de contagio de la enfermedad T es censurado por la izquierda por C entonces $X = C$, y $A = 2$. Finalmente, en el caso iii) T es censurado por la derecha por C entonces $X = C$ y $A = 1$. Es decir, se observan variables de la forma (X, A) con $X \geq 0$ y $A \in \{0, 1, 2\}$. Notése que si no hay observaciones del tipo ii) se tiene un caso clásico de censura por la derecha y si no hay observaciones del tipo i) se tienen datos de estado actual.

En la siguiente sección, se describirá el mecanismo de censura anterior introduciendo variables latentes y se harán suposiciones acerca de sus distribuciones. A continuación, en la Sección 3.3, se propone ajustar un modelo de regresión de Cox para datos con la estructura antes descrita, obteniendo los estimadores de los efectos de las covariables y la función de riesgo base para el modelo. En la Sección 3.4, se demostrarán las propiedades asintóticas de los estimadores propuestos en la Sección 3.3. Finalmente, en las Secciones 3.6 y 3.5 se realizará un estudio de simulación para analizar empíricamente las propiedades de los estimadores y se ajustará dicho modelo a datos reales sobre un estudio de VIH en mujeres, respectivamente.

3.2. Modelo Latente para la Estructura de Datos

Sea T el tiempo de supervivencia de interés, C el tiempo de censura y Δ una variable aleatoria Bernoulli. Se asume que las variables T , C y Δ son variables independientes latentes. Las variables observadas son de la forma (X, A) , con $X \geq 0$ y $A \in \{0, 1, 2\}$. Las variables observadas se pueden definir en función de las latentes como sigue

$$\begin{aligned} X &= T, \quad A = 0 \quad \text{si } 0 \leq T \leq C \text{ y } \Delta = 1, \\ X &= C, \quad A = 1 \quad \text{si } 0 \leq C \leq T, \\ X &= C, \quad A = 2 \quad \text{si } 0 \leq T \leq C \text{ y } \Delta = 0. \end{aligned} \tag{3.2.1}$$

Además, se pueden escribir como

$$X = \min(T, C) + (1 - \Delta) \max(C - T, 0)$$

y

$$A = 2(1 - \Delta)\mathbb{1}_{\{T \leq C\}} + \mathbb{1}_{\{C < T\}}.$$

Obsérvese que si $\Delta = 1$ casi seguro se tienen datos con censura por la derecha y si $\Delta = 0$ casi seguro se tienen datos de estado actual.

Sea $F_T(\cdot)$ y $F_C(\cdot)$ las distribuciones de T y C , respectivamente y $p = P(\Delta = 1)$. Se definen las subdistribuciones

$$H_k(B) = P(X \in B, A = k), \quad k = 0, 1, 2,$$

donde B es un conjunto de Borel en $[0, +\infty]$. De este modo, la distribución de X está dada por $H(\cdot) = H_0(\cdot) + H_1(\cdot) + H_2(\cdot)$.

Se puede definir una función Φ entre la distribución de las variables latentes y la distribución de los tiempos observados de tal modo que

$$(H_0, H_1, H_2) = \Phi(F_T, F_C, p)$$

se define como

$$\begin{aligned} H_0(dt) &= pF_C([t, \infty])F_T(dt), \\ H_1(dt) &= F_T((t, \infty])F_C(dt), \\ H_2(dt) &= (1 - p)F_T([0, t])F_C(dt). \end{aligned} \tag{3.2.2}$$

Nótese que si $p = 1$ (ó $p = 0$) las ecuaciones anteriores se reducen al modelo clásico de datos con censura por la derecha (ó datos de estado actual).

3.3. Estimadores del Modelo de Cox

Como se ha mencionado antes, el objetivo de este capítulo es ajustar un modelo de regresión de Cox a datos doblemente censurados. El modelo de Cox viene dado por

$$\lambda(t|\mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z}), \tag{3.3.3}$$

donde $\lambda_0(\cdot)$ es la función de riesgo base, $\mathbf{z} = (z_1, \dots, z_d)$ es el vector de covariables y $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)$ es el vector de parámetros de regresión. En este caso, se considera $\lambda_0(\cdot)$ una función arbitraria desconocida, es decir, un modelo semiparamétrico, por tanto se debe obtener una estimación del vector de parámetros $\boldsymbol{\beta}$ y una estimación de la función de riesgo base acumulativa $\Lambda_0(\cdot)$.

Se asume que el tiempo de fallo t es generado bajo el modelo (3.3.3) dado \mathbf{z} y que es condicionalmente independiente del tiempo de censura c y que éste no depende de \mathbf{z} . Entonces el modelo de variables latentes (3.2.2) es

$$\begin{aligned} H_0(dt|\mathbf{z}) &= pF_C([t, +\infty])F_T(dt|\mathbf{z}) \\ H_1(dt|\mathbf{z}) &= F_T((t, +\infty]|\mathbf{z})F_C(dt) \\ H_2(dt|\mathbf{z}) &= (1 - p)F_T([0, t]|\mathbf{z})F_C(dt). \end{aligned} \tag{3.3.4}$$

En primer lugar, se requiere un estimador de la función de riesgo base acumulativa $\Lambda_0(\cdot)$, para ello se sigue el siguiente camino. Si se integra por partes sobre

$[t, +\infty]$ la primera ecuación de (3.3.4) se obtiene

$$\begin{aligned}
 H_0([t, +\infty]|\mathbf{z}) &= \int_t^\infty H_0(dt|\mathbf{z}) \\
 &= p \int_t^\infty F_C([t, +\infty])F_T(dt|\mathbf{z}) \\
 &= p \left[F_C([t, +\infty])F_T([t, +\infty]) - \int_t^\infty F_T([t, +\infty]|\mathbf{z})F_C(dt) \right] \\
 &= pF_C([t, +\infty])F_T([t, +\infty]) - p \int_t^\infty H_1(dt|\mathbf{z}) \\
 &= pF_C([t, +\infty])F_T([t, +\infty]) - pH_1([t, +\infty]|\mathbf{z}),
 \end{aligned}$$

de donde se deduce que

$$H_0([t, +\infty]|\mathbf{z}) + pH_1([t, +\infty]|\mathbf{z}) = pF_T([t, +\infty]|\mathbf{z})F_C([t, +\infty]). \quad (3.3.5)$$

Además, si $t = 0$ se obtiene

$$H_0([0, +\infty]|\mathbf{z}) + pH_1([0, +\infty]|\mathbf{z}) = p$$

ya que $F_T([0, +\infty]|\mathbf{z}) = 1$ y $F_C([0, +\infty]|\mathbf{z}) = 1$. Por tanto,

$$H_0([0, +\infty]|\mathbf{z}) = p(1 - H_1([0, +\infty]|\mathbf{z})). \quad (3.3.6)$$

Nótese que

$$1 = H_0([0, +\infty]|\mathbf{z}) + H_1([0, +\infty]|\mathbf{z}) + H_2([0, +\infty]|\mathbf{z}),$$

entonces

$$1 - H_1([0, +\infty]|\mathbf{z}) = H_0([0, +\infty]|\mathbf{z}) + H_2([0, +\infty]|\mathbf{z}).$$

Por tanto, la ecuación (3.3.6) se puede escribir como

$$H_0([0, +\infty]|\mathbf{z}) = p(H_0([0, +\infty]|\mathbf{z}) + H_2([0, +\infty]|\mathbf{z})). \quad (3.3.7)$$

Por otro lado, por la expresión (1.1.7), se sabe que la medida de riesgo asociada a $F_T(\cdot|\mathbf{z})$ es

$$\begin{aligned}\Lambda(dt|\mathbf{z}) &= \frac{F_T(dt|\mathbf{z})}{F_T([t, +\infty]|\mathbf{z})} \\ &= \frac{pF_C([t, +\infty])F_T(dt|\mathbf{z})}{pF_C([t, +\infty])F_T([t, +\infty]|\mathbf{z})} \\ &= \frac{H_0(dt|\mathbf{z})}{H_0([t, +\infty]|\mathbf{z}) + pH_1([t, +\infty]|\mathbf{z})},\end{aligned}\tag{3.3.8}$$

teniendo en cuenta la primera ecuación de (3.3.4) y la ecuación (3.3.5).

Obsérvese que la expresión anterior es válida si $t \in I$, donde $I = \{t \in \mathbb{R}^+, \Lambda_0 < +\infty \text{ y } F_C([t, +\infty]) > 0\}$ y para $t \in I^c$ no se tiene información disponible. También se requiere que $p \neq 0$, en el caso $p = 0$ se debe aplicar el modelo para datos de estado actual.

Para el modelo de Cox (3.3.3) de la ecuación (3.3.8) se tiene

$$\begin{aligned}\frac{H_0(dt|\mathbf{z})}{H_0([t, +\infty]|\mathbf{z}) + pH_1([t, +\infty]|\mathbf{z})} &= \lambda(t|\mathbf{z})dt \\ &= \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z})dt \\ &= \exp(\boldsymbol{\beta}'\mathbf{z})\lambda_0(t)dt\end{aligned}\tag{3.3.9}$$

entonces

$$\lambda_0(t)dt = \frac{H_0(dt|\mathbf{z})}{\exp(\boldsymbol{\beta}'\mathbf{z}) [H_0([t, +\infty]|\mathbf{z}) + pH_1([t, +\infty]|\mathbf{z})]}.$$

Finalmente, se tiene

$$\begin{aligned}\Lambda_0(t) &= \int_0^t \frac{H_0(ds|\mathbf{z})}{\exp(\boldsymbol{\beta}'\mathbf{z}) [H_0([s, +\infty]|\mathbf{z}) + pH_1([s, +\infty]|\mathbf{z})]} \\ &= \int_0^t \frac{H_0(ds|\mathbf{z})}{\exp(\boldsymbol{\beta}'\mathbf{z}) H_0([s, +\infty]|\mathbf{z}) + p \exp(\boldsymbol{\beta}'\mathbf{z}) H_1([s, +\infty]|\mathbf{z})}.\end{aligned}\tag{3.3.10}$$

Sea una muestra de vectores iid de la forma (X_i, A_i, \mathbf{Z}_i) con $X_i = \min(T_i, C_i) + (1 - \Delta_i) \max(C_i - T_i, 0)$, $A_i = 2(1 - \Delta_i)\mathbb{1}_{\{T_i \leq C_i\}} + \mathbb{1}_{\{C_i < T_i\}}$ y \mathbf{Z}_i el vector de covariables para el individuo i con $i = 1, \dots, n$.

De la ecuación (3.3.7) y sumando sobre i en ambos lados de la ecuación se tiene que p viene dado por

$$\begin{aligned} p &= \frac{\sum_{i=1}^n H_0([0, +\infty]|\mathbf{z}_i)}{\sum_{i=1}^n (H_0([0, +\infty]|\mathbf{z}_i) + H_2([0, +\infty]|\mathbf{z}_i))} \\ &= \frac{\sum_{i=1}^n H_0([0, +\infty]|\mathbf{z}_i)}{\sum_{i=1}^n H_0([0, +\infty]|\mathbf{z}_i) + \sum_{i=1}^n H_2([0, +\infty]|\mathbf{z}_i)}. \end{aligned}$$

Si se define

$$N_k = \sum_{i=1}^n \mathbb{1}_{\{A_i=k\}}, \quad (3.3.11)$$

el estimador de p es

$$\hat{p} = \frac{N_0}{N_0 + N_2}.$$

Además, si se define

$$S_k^{(q)}(t|\boldsymbol{\beta}) = \sum_{i=1}^n \exp(\boldsymbol{\beta}'\mathbf{z}_i) \mathbf{z}_i^{\otimes q} \mathbb{1}_{\{X_i \geq t, A_i=k\}} \quad (3.3.12)$$

y

$$N_k(t) = \sum_{i=1}^n \mathbb{1}_{\{X_i \leq t, A_i=k\}}$$

donde $\mathbf{z}^{\otimes 0} = 1$, $\mathbf{z}^{\otimes 1} = \mathbf{z}$, $\mathbf{z}^{\otimes 2} = \mathbf{z}\mathbf{z}'$ con \mathbf{z} un vector columna y se considera la expresión (3.3.10), se puede estimar $\Lambda_0(\cdot)$ por

$$\hat{\Lambda}_0(t) = \int_0^t \frac{N_0(ds)}{S_0^{(0)}(s|\boldsymbol{\beta}) + \hat{p}S_1^{(0)}(s|\boldsymbol{\beta})}, \quad (3.3.13)$$

por tanto,

$$\hat{\lambda}_0(t) = \frac{N_0(t)}{S_0^{(0)}(t|\boldsymbol{\beta}) + \hat{p}S_1^{(0)}(t|\boldsymbol{\beta})}. \quad (3.3.14)$$

Una vez estimadas las funciones $\Lambda_0(\cdot)$ y $\lambda_0(\cdot)$ se puede obtener una estimación de la función de supervivencia $\hat{S}(\cdot)$ y de la función de densidad $\hat{f}(\cdot)$ del modelo de Cox (3.3.3) como se describe en la Sección 1.2. Si se utilizan dichas expresiones

en la función de verosimilitud dada por (1.1.8) se tiene que la función de log-verosimilitud está en función de los parámetros β , esto es

$$\begin{aligned}
 \mathcal{L}_n(\beta) &= \prod_{i=1}^n \left[\hat{f}_i(x_i) \right]^{\Delta N_{0i}} \left[\hat{S}_i(x_i) \right]^{\Delta N_{1i}} \left[1 - \hat{S}_i(x_i) \right]^{\Delta N_{2i}} \\
 &= \prod_{i=1}^n \left[\hat{\lambda}_i(x_i) \hat{S}_i(x_i) \right]^{\Delta N_{0i}} \left[\hat{S}_i(x_i) \right]^{\Delta N_{1i}} \left[1 - \hat{S}_i(x_i) \right]^{\Delta N_{2i}} \\
 &= \prod_{i=1}^n \left\{ \left[\frac{\exp(\beta' \mathbf{z}_i)}{S_0^{(0)}(x_i|\beta) + \hat{p}S_1^{(0)}(x_i|\beta)} \right]^{\Delta N_{0i}} \right. \\
 &\quad \left[1 - \exp \left(-\exp(\beta' \mathbf{z}_i) \int_0^{x_i} \frac{N_0(ds)}{S_0^{(0)}(s|\beta) + \hat{p}S_1^{(0)}(s|\beta)} \right) \right]^{\Delta N_{2i}} \\
 &\quad \left. \left[\exp \left(-\exp(\beta' \mathbf{z}_i) \int_0^{x_i} \frac{N_0(ds)}{S_0^{(0)}(s|\beta) + \hat{p}S_1^{(0)}(s|\beta)} \right) \right]^{\Delta N_{0i} + \Delta N_{1i}} \right\} \\
 &= \prod_{i=1}^n \left\{ \left[\frac{\exp(\beta' \mathbf{z}_i)}{S_0^{(0)}(x_i|\beta) + \hat{p}S_1^{(0)}(x_i|\beta)} \right]^{\Delta N_{0i}} \right. \\
 &\quad \left[1 - \exp \left(-\exp(\beta' \mathbf{z}_i) \int_0^{x_i} \frac{N_0(ds)}{S_0^{(0)}(s|\beta) + \hat{p}S_1^{(0)}(s|\beta)} \right) \right]^{\Delta N_{2i}} \\
 &\quad \left. \left[\exp \left(-\int_0^{x_i} \frac{S_0^{(0)}(s|\beta) + S_1^{(0)}(s|\beta)}{S_0^{(0)}(s|\beta) + \hat{p}S_1^{(0)}(s|\beta)} N_0(ds) \right) \right] \right\}
 \end{aligned}$$

con $\Delta N_{ki} = \Delta N_k(x_i)$, para $i = 1, \dots, n$ y $k = 0, 1, 2$.

Tomando el logaritmo de la función anterior, se tiene que la función de log-

verosimilitud viene dada por

$$\begin{aligned}
 \ell_n(\boldsymbol{\beta}) &= \log \mathcal{L}_n(\boldsymbol{\beta}) \\
 &= \sum_{i=1}^n \left(\Delta N_{0i} \log \left(\frac{\exp(\boldsymbol{\beta}' \mathbf{z}_i)}{E^{(0)}(x_i | \boldsymbol{\beta}, \hat{p})} \right) - \int_0^{x_i} \frac{E^{(0)}(s | \boldsymbol{\beta}, 1)}{E^{(0)}(s | \boldsymbol{\beta}, \hat{p})} N_0(ds) \right. \\
 &\quad \left. + \Delta N_{2i} \log \left(1 - \exp \left(- \exp(\boldsymbol{\beta}' \mathbf{z}_i) \int_0^{x_i} \frac{N_0(ds)}{E^{(0)}(s | \boldsymbol{\beta}, \hat{p})} \right) \right) \right) \\
 &= \sum_{i=1}^n \left(\Delta N_{0i} \boldsymbol{\beta}' \mathbf{z}_i - \Delta N_{0i} \log (E^{(0)}(x_i | \boldsymbol{\beta}, \hat{p})) - \int_0^{x_i} \frac{E^{(0)}(s | \boldsymbol{\beta}, 1)}{E^{(0)}(s | \boldsymbol{\beta}, \hat{p})} N_0(ds) \right. \\
 &\quad \left. + \Delta N_{2i} \log \left(1 - \exp \left(- \exp(\boldsymbol{\beta}' \mathbf{z}_i) \int_0^{x_i} \frac{N_0(ds)}{E^{(0)}(s | \boldsymbol{\beta}, \hat{p})} \right) \right) \right) \quad (3.3.15)
 \end{aligned}$$

donde $E^{(k)}(s | \boldsymbol{\beta}, p) = S_0^{(k)}(s | \boldsymbol{\beta}) + p S_1^{(k)}(s | \boldsymbol{\beta})$.

Por tanto, el estimador de máxima verosimilitud “profile” del vector de parámetros desconocidos del modelo de Cox (3.3.3) para datos doblemente censurados como los descritos en (3.2.1) viene dado por

$$\tilde{\boldsymbol{\beta}}_n = \arg \max_{\boldsymbol{\beta}} \ell_n(\boldsymbol{\beta}).$$

Puesto que no existe una fórmula explícita, el problema de optimización anterior se resuelve con métodos numéricos del tipo Newton–Raphson. En el Apéndice I de este trabajo se ha escrito la expresión de la derivada parcial de la función (3.3.15), con la finalidad de obtener su función gradiente que es utilizada en los métodos numéricos de optimización. En la práctica se usará la función de optimización `optim` del Software estadístico **R**. En dicha función se utiliza el método de maximización BFGS.

3.4. Propiedades Asintóticas de los Estimadores de los Parámetros

Como en el capítulo anterior, para demostrar la normalidad asintótica de los estimadores de los parámetros del modelo se aplicará el Teorema principal de Murphy y van der Vaart (2000). Esto es, es necesario verificar que se cumplen las condiciones (1.3.31)–(1.3.34) y las hipótesis del Teorema 1.3.10, para poder aplicar dicho teorema.

Condiciones 3.4.1. *Para demostrar las propiedades asintóticas de los estimadores $\hat{\beta}$ del modelo (3.3.3) se harán las siguientes suposiciones:*

- i) *el tiempo observado está dentro del intervalo $[\sigma, \tau]$ con $0 < \sigma < \tau < \infty$,*
- ii) *el vector de covariables \mathbf{Z} es acotado,*
- iii) *el verdadero valor β_0 del parámetro del modelo se encuentra en el interior de un compacto conocido $\Theta \subset \mathbb{R}^d$ y la verdadera función de riesgo base $\Lambda_0^0(\cdot)$ está en un convexo \mathcal{B} del espacio de funciones,*
- iv) *$P_0(X \geq \tau) = E_0[\mathbb{1}_{\{X \geq \tau\}}] > 0$, donde $E_0[f]$ denota el valor esperado de f bajo la verdadera función de distribución P_0 ,*
- v) *la función de riesgo base acumulativa $\Lambda_0(\cdot)$ es uniformemente acotada y de variación uniformemente acotada. Además, $\Lambda_0(\cdot)$ es estrictamente creciente en $[\sigma, \tau]$ con $\Lambda_0(\sigma) > 0$ y $\Lambda_0(\tau) < M < \infty$ y es continuamente diferenciable,*
y
- vi) *la matriz de información eficiente $\tilde{\mathbf{I}}_0$ es invertible.*

Bajo el modelo latente (3.3.4), se observan n vectores iid $Y_i = (X_i, A_i, \mathbf{Z}_i) \sim P_{\beta, \Lambda_0} \in \mathcal{P}$, con $i = 1, \dots, n$, donde \mathcal{P} es el modelo semiparamétrico dado por

$$\mathcal{P} = \{P_{\beta, \Lambda_0} : \beta \in \mathbb{R}^d \text{ es el vector de parámetros del modelo de Cox (3.3.3) y } \Lambda_0(\cdot) \text{ es la función de riesgo base acumulativa}\},$$

con $p_{\beta, \Lambda_0}(\cdot)$ la función de densidad correspondiente a $P_{\beta, \Lambda_0} \in \mathcal{P}$.

Además, la función

$$\ell_n(\beta, \Lambda_0) = \sum_{i=1}^n \ell(\beta, \Lambda_0)(\mathbf{y}_i) = \sum_{i=1}^n \log p_{\beta, \Lambda_0}(\mathbf{y}_i) \quad (3.4.16)$$

es la función de log-verosimilitud total para la muestra de n observaciones. Por tanto, la función de log-verosimilitud “profile” para β viene dada por

$$p\ell_n(\beta) = \sup_{\Lambda_0} \ell_n(\beta, \Lambda_0). \quad (3.4.17)$$

Dado el mecanismo de censura de datos doblemente censurados, se puede asumir que la función de densidad $p_{\beta, \Lambda_0}(\cdot)$ tiene la forma

$$\begin{aligned} p_{\beta, \Lambda_0}(\mathbf{y}) &= [\exp(\beta' \mathbf{z}) \lambda_0(x) S_{T|\mathbf{Z}}(x|\mathbf{z}) S_{C|\mathbf{Z}}(x|\mathbf{z})]^{n_0} [S_{T|\mathbf{Z}}(x|\mathbf{z}) p_{C|\mathbf{Z}}(x|\mathbf{z})]^{n_1} \\ &\quad [(1 - S_{T|\mathbf{Z}}(x|\mathbf{z})) p_{C|\mathbf{Z}}(x|\mathbf{z})]^{n_2} p_{\mathbf{Z}}(\mathbf{z}) \\ &= [\exp(\beta' \mathbf{z}) \lambda_0(x) \exp(-\Lambda_0(x) \exp(\beta' \mathbf{z})) S_{C|\mathbf{Z}}(x|\mathbf{z})]^{n_0} \\ &\quad [\exp(-\Lambda_0(x) \exp(\beta' \mathbf{z})) p_{C|\mathbf{Z}}(x|\mathbf{z})]^{n_1} \\ &\quad [(1 - \exp(-\Lambda_0(x) \exp(\beta' \mathbf{z}))) p_{C|\mathbf{Z}}(x|\mathbf{z})]^{n_2} p_{\mathbf{Z}}(\mathbf{z}) \end{aligned}$$

donde $n_k = \mathbb{1}_{\{A=k\}}$. Se observa que la función de densidad $p_{\beta, \Lambda_0}(\mathbf{y})$ es proporcional a

$$[\exp(\beta' \mathbf{z}) \lambda_0(x)]^{n_0} [\exp(-\Lambda_0(x) \exp(\beta' \mathbf{z}))]^{n_2} [1 - \exp(-\Lambda_0(x) \exp(\beta' \mathbf{z}))]^{n_2}$$

con $n_{-k} = \mathbb{1}_{\{A \neq k\}}$. Por tanto, la función de log-verosimilitud para \mathbf{y} se puede escribir como:

$$\begin{aligned} \ell(\beta, \Lambda_0)(\mathbf{y}) &= n_0 [\beta' \mathbf{z} + \log \lambda_0(x)] - n_{-2} [\Lambda_0(x) \exp(\beta' \mathbf{z})] \\ &\quad + n_2 \log[1 - \exp(-\Lambda_0(x) \exp(\beta' \mathbf{z}))] + cte. \end{aligned} \quad (3.4.18)$$

Lo primero que se debe hacer es encontrar la función “score” eficiente del modelo. Para obtener la función “score” eficiente dada por la expresión (1.3.30), en primer lugar es necesario tener la función “score” con respecto a β de la ecuación (3.4.18) que viene dada por

$$\begin{aligned}
 \frac{\partial}{\partial \beta} \ell(\beta, \Lambda_0)(\mathbf{y}) &= n_0 \mathbf{z} - n_{-2} [\Lambda_0(x) \exp(\beta' \mathbf{z}) \mathbf{z}] \\
 &\quad + n_2 \frac{\exp(-\Lambda_0(x) \exp(\beta' \mathbf{z})) \Lambda_0(x) \exp(\beta' \mathbf{z}) \mathbf{z}}{1 - \exp(-\Lambda_0(x) \exp(\beta' \mathbf{z}))} \\
 &= n_0 \mathbf{z} - n_{-2} [\exp(\beta' \mathbf{z}) \Lambda_0(x) \mathbf{z}] + n_2 \frac{\exp(\beta' \mathbf{z}) \mathbf{z} \Lambda_0(x)}{\exp(\Lambda_0(x) \exp(\beta' \mathbf{z})) - 1} \\
 &= n_0 \mathbf{z} + \left(\frac{n_2}{\exp(\Lambda_0(x) \exp(\beta' \mathbf{z})) - 1} - n_{-2} \right) \mathbf{z} \Lambda_0(x) \exp(\beta' \mathbf{z}) \\
 &= n_0 \mathbf{z} + Q_{\beta, \Lambda_0}(\mathbf{y}) \mathbf{z} \Lambda_0(x) \\
 &= \dot{\ell}(\beta, \Lambda_0)(\mathbf{y}),
 \end{aligned}$$

con $Q_{\beta, \Lambda_0}(\mathbf{y}) = \left(\frac{n_2}{\exp(\Lambda_0(x) \exp(\beta' \mathbf{z})) - 1} - n_{-2} \right) \exp(\beta' \mathbf{z})$.

Para obtener la segunda parte de la función “score” eficiente (1.3.30), se define la curva

$$\Lambda_{\mathbf{t}}(x) = \int_{[\sigma, x]} (1 + \mathbf{t}' h(s)) d\Lambda_0(s), \quad (3.4.19)$$

para $\Lambda_0(\cdot)$ dada y una función acotada $h : [\sigma, \tau] \mapsto \mathbb{R}^d$. Si $h(\cdot)$ es una dirección menos favorable en el verdadero valor $\Lambda_0^0(\cdot)$, la función anterior define un submodelo menos favorable que pasa a través del verdadero modelo en $\mathbf{t} = \mathbf{0}$ para $\Lambda_0^0(\cdot)$ dada.

Ahora, se sustituye el submodelo (3.4.19) en la expresión (3.4.18) y se obtiene

$$\begin{aligned}
 \ell(\beta, \Lambda_{\mathbf{t}})(\mathbf{y}) &= n_0 [\beta' \mathbf{z} + \log((1 + \mathbf{t}' h(x)) \lambda_0(x))] \\
 &\quad - n_{-2} \left[\int_{[\sigma, x]} (1 + \mathbf{t}' h(s)) d\Lambda_0(s) \exp(\beta' \mathbf{z}) \right] \\
 &\quad + n_2 \log \left[1 - \exp \left(- \int_{[\sigma, x]} (1 + \mathbf{t}' h(s)) d\Lambda_0(s) \exp(\beta' \mathbf{z}) \right) \right],
 \end{aligned}$$

se diferencia la expresión anterior con respecto a \mathbf{t} y se evalúa en $\mathbf{t} = \mathbf{0}$ ya que el

submodelo pasa a través del modelo en ese punto, así se obtiene:

$$\begin{aligned}
 & \frac{\partial}{\partial \mathbf{t}} \ell(\boldsymbol{\beta}, \Lambda_{\mathbf{t}})(\mathbf{y})|_{\mathbf{t}=\mathbf{0}} \\
 &= n_0 \frac{h(x)\lambda_0(x)}{(1 + \mathbf{t}'h(x))\lambda_0(x)} \Big|_{\mathbf{t}=\mathbf{0}} - n_{-2} \int_{[\sigma, x]} h(s) d\Lambda_0(s) \exp(\boldsymbol{\beta}'\mathbf{z}) \\
 & \quad + n_2 \frac{\exp\left(-\int_{[\sigma, x]} (1 + \mathbf{t}'h(s)) d\Lambda_0(s) \exp(\boldsymbol{\beta}'\mathbf{z})\right) \int_{[\sigma, x]} h(s) d\Lambda_0(s) \exp(\boldsymbol{\beta}'\mathbf{z})}{1 - \exp\left(-\int_{[\sigma, x]} (1 + \mathbf{t}'h(s)) d\Lambda_0(s) \exp(\boldsymbol{\beta}'\mathbf{z})\right)} \Big|_{\mathbf{t}=\mathbf{0}} \\
 &= n_0 h(x) - n_{-2} \exp(\boldsymbol{\beta}'\mathbf{z}) \int_{[\sigma, x]} h(s) d\Lambda_0(s) \\
 & \quad + n_2 \frac{\exp(-\exp(\boldsymbol{\beta}'\mathbf{z})\Lambda_0(x)) \exp(\boldsymbol{\beta}'\mathbf{z}) \int_{[\sigma, x]} h(s) d\Lambda_0(s)}{1 - \exp(-\exp(\boldsymbol{\beta}'\mathbf{z})\Lambda_0(x))} \\
 &= n_0 h(x) - n_{-2} \exp(\boldsymbol{\beta}'\mathbf{z}) \int_{[\sigma, x]} h(s) d\Lambda_0(s) + n_2 \frac{\exp(\boldsymbol{\beta}'\mathbf{z}) \int_{[\sigma, x]} h(s) d\Lambda_0(s)}{\exp(\Lambda_0(x) \exp(\boldsymbol{\beta}'\mathbf{z})) - 1} \\
 &= n_0 h(x) + \left(\frac{n_2}{\exp(\Lambda_0(x) \exp(\boldsymbol{\beta}'\mathbf{z})) - 1} - n_{-2} \right) \exp(\boldsymbol{\beta}'\mathbf{z}) \int_{[\sigma, x]} h(s) d\Lambda_0(s) \\
 &= n_0 h(x) + Q_{\boldsymbol{\beta}, \Lambda_0}(\mathbf{y}) \int_{[\sigma, x]} h(s) d\Lambda_0(s) \\
 &= A_{\boldsymbol{\beta}, \Lambda_0} h(\mathbf{y}).
 \end{aligned}$$

Puesto que, en la expresión (1.3.30), $\Pi_{\boldsymbol{\beta}, \Lambda_0}$ es la proyección ortogonal sobre la clausura del espacio lineal generado por $\mathcal{P}_{\boldsymbol{\beta}, \Lambda_0}$ se tiene que la función “score” eficiente del modelo es

$$\tilde{\ell}(\boldsymbol{\beta}, \Lambda_0)(\mathbf{y}) = \dot{\ell}(\boldsymbol{\beta}, \Lambda_0)(\mathbf{y}) - A_{\boldsymbol{\beta}, \Lambda_0} h_{\boldsymbol{\beta}, \Lambda_0}(\mathbf{y}),$$

con $h_{\boldsymbol{\beta}, \Lambda_0}(\mathbf{y}) = \frac{A_{\boldsymbol{\beta}, \Lambda_0}^* \dot{\ell}(\boldsymbol{\beta}, \Lambda_0)(\mathbf{y})}{A_{\boldsymbol{\beta}, \Lambda_0}^* A_{\boldsymbol{\beta}, \Lambda_0}(x)}$, donde $A_{\boldsymbol{\beta}, \Lambda_0}^*$ es el operador adjunto de $A_{\boldsymbol{\beta}, \Lambda_0}$. Para obtener las expresiones $A_{\boldsymbol{\beta}, \Lambda_0}^* A_{\boldsymbol{\beta}, \Lambda_0}$ y $A_{\boldsymbol{\beta}, \Lambda_0}^* \dot{\ell}(\boldsymbol{\beta}, \Lambda_0)$ se usan las igualdades (2.6.21) y (2.6.22), respectivamente. Esto es,

$$\begin{aligned}
 (A_{\boldsymbol{\beta}, \Lambda_0} g(\mathbf{y}))' (A_{\boldsymbol{\beta}, \Lambda_0} h(\mathbf{y})) &= n_0 g(x)' h(x) + Q_{\boldsymbol{\beta}, \Lambda_0}^2(\mathbf{y}) \left(\int_{\sigma}^x g(s) d\Lambda_0(s) \right)' \int_{\sigma}^x h(s) d\Lambda_0(s) \\
 & \quad + n_0 Q_{\boldsymbol{\beta}, \Lambda_0}(\mathbf{y}) \left[g(x)' \int_{\sigma}^x h(s) d\Lambda_0(s) + h(x)' \int_{\sigma}^x g(s) d\Lambda_0(s) \right],
 \end{aligned}$$

se define

$$a(x) = g(x)' \int_{\sigma}^x h(s) d\Lambda_0(s) + h(x)' \int_{\sigma}^x g(s) d\Lambda_0(s)$$

entonces, integrando por partes como en el Apendice II, se tiene

$$\begin{aligned} (A_{\beta, \Lambda_0} g(\mathbf{y}))' (A_{\beta, \Lambda_0} h(\mathbf{y})) &= n_0 g(x)' h(x) + Q_{\beta, \Lambda_0}^2(\mathbf{y}) \int_{\sigma}^x a(s) d\Lambda_0(s) \\ &\quad + n_0 Q_{\beta, \Lambda_0}(\mathbf{y}) a(x). \end{aligned}$$

Tomando la esperanza se tiene

$$\begin{aligned} P_{\beta, \Lambda_0} [n_0 g(x)' h(x)] &= \int_{\sigma}^{\tau} \int_{\mathbb{R}^d} g(x)' h(x) \exp(\beta' \mathbf{z}) S_{T|\mathbf{Z}}(x|\mathbf{z}) S_{C|\mathbf{Z}}(x|\mathbf{z}) p(\mathbf{z}) d\mathbf{z} d\Lambda_0(x) \\ &= \int_{\sigma}^{\tau} g(x)' \left[\int_{\mathbb{R}^d} \exp(\beta' \mathbf{z}) E[\mathbb{1}_{\{X \geq x\}} | \mathbf{z}] p(\mathbf{z}) d\mathbf{z} \right] h(x) d\Lambda_0(x) \\ &= \int_{\sigma}^{\tau} g(x)' E[\exp(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}}] h(x) d\Lambda_0(x), \end{aligned} \tag{3.4.20}$$

y

$$\begin{aligned} P_{\beta, \Lambda_0} [n_0 Q_{\beta, \Lambda_0}(\mathbf{y}) a(x)] &= P_{\beta, \Lambda_0} [-n_0 \exp(\beta' \mathbf{z}) a(x)] \\ &= - \int_{\sigma}^{\tau} \int_{\mathbb{R}^d} a(x) \exp^2(\beta' \mathbf{z}) S_{T|\mathbf{Z}}(x|\mathbf{z}) S_{C|\mathbf{Z}}(x|\mathbf{z}) p(\mathbf{z}) d\mathbf{z} d\Lambda_0(x) \\ &= - \int_{\sigma}^{\tau} a(x) E[\exp^2(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}}] d\Lambda_0(s) d\Lambda_0(x) \end{aligned}$$

además,

$$P_{\beta, \Lambda_0} \left[Q_{\beta, \Lambda_0}^2(\mathbf{y}) \int_{\sigma}^x a(s) d\Lambda_0(s) \right] = \int_{\sigma}^{\tau} a(s) E[Q_{\beta, \Lambda_0}^2(\mathbf{y}) \mathbb{1}_{\{X \geq x\}}] d\Lambda_0(s).$$

Por tanto,

$$\begin{aligned} P_{\beta, \Lambda_0} [(A_{\beta, \Lambda_0} g)(A_{\beta, \Lambda_0} h)] &= \int_{\sigma}^{\tau} (g(x)' E[\exp(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}}] h(x) \\ &\quad + a(x) [E[Q_{\beta, \Lambda_0}^2(\mathbf{y}) \mathbb{1}_{\{X \geq x\}}] - E[\exp^2(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}}]]) d\Lambda_0(x) \end{aligned}$$

y por la expresión (2.6.21) se tiene

$$\begin{aligned} g(x)' A_{\beta, \Lambda_0}^* A_{\beta, \Lambda} h(\mathbf{y}) &= g(x)' E[\exp(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}}] h(x) \\ &\quad + a(x) [E[Q_{\beta, \Lambda_0}^2(\mathbf{y}) \mathbb{1}_{\{X \geq x\}}] - E[\exp^2(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}}]], \end{aligned}$$

de donde

$$\begin{aligned} & A_{\beta, \Lambda_0}^* A_{\beta, \Lambda}(x) \\ &= E \left[\exp(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}} \right] + 2\Lambda_0(x) \left[E \left[Q_{\beta, \Lambda_0}^2(\mathbf{y}) \mathbb{1}_{\{X \geq x\}} \right] - E \left[\exp^2(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}} \right] \right]. \end{aligned}$$

Por otro lado, análogamente se define

$$a(x) = \int_{\sigma}^x h(s) d\Lambda_0(s) + h(x) \Lambda_0(x)$$

e integrando por partes como en el Apéndice II se tiene

$$\left(\dot{\ell}(\beta, \Lambda_0)(\mathbf{y}) \right)' (A_{\beta, \Lambda_0} h(\mathbf{y})) = n_0 \mathbf{z}' h(x) - n_0 \mathbf{z}' \exp(\beta' \mathbf{z}) a(x) + \mathbf{z}' Q_{\beta, \Lambda_0}^2(\mathbf{y}) \int_{\sigma}^x a(s) d\Lambda_0(s).$$

Tomando la esperanza

$$\begin{aligned} & P_{\beta, \Lambda_0} \left[(\dot{\ell}(\beta, \Lambda_0)(A_{\beta, \Lambda_0} h)) \right] \\ &= \int_{\sigma}^{\tau} \left(E \left[\mathbf{z} \exp(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}} \right]' h(x) + \left[E \left[\mathbf{z} Q_{\beta, \Lambda_0}^2(\mathbf{y}) \mathbb{1}_{\{X \geq x\}} \right] \right. \right. \\ &\quad \left. \left. - E \left[\mathbf{z} \exp^2(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}} \right] \right]' a(x) \right) d\Lambda_0(x) \end{aligned}$$

y por la expresión (2.6.22) se tiene

$$\begin{aligned} A_{\beta, \Lambda_0}^* \dot{\ell}(\beta, \Lambda)(\mathbf{y}) &= E \left[\mathbf{z} \exp(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}} \right] \\ &\quad + 2\Lambda_0(x) \left[E \left[\mathbf{z} Q_{\beta, \Lambda_0}^2(\mathbf{y}) \mathbb{1}_{\{X \geq x\}} \right] - E \left[\mathbf{z} \exp^2(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}} \right] \right], \end{aligned}$$

Si $A_{\beta, \Lambda_0}^* A_{\beta, \Lambda_0}$ es un operador continuamente invertible en el espacio de funciones acotadas de variación acotada y $A_{\beta, \Lambda_0}^* \dot{\ell}(\beta, \Lambda_0)$ es un vector de funciones acotadas de variación acotada, la dirección menos favorable es

$$h_{\beta, \Lambda_0}(x) = \frac{A_{\beta, \Lambda_0}^* \dot{\ell}(\beta, \Lambda_0)(x)}{A_{\beta, \Lambda_0}^* A_{\beta, \Lambda_0}(x)}.$$

Entonces, se tiene que la función “score” eficiente está dada por

$$\begin{aligned} \tilde{\ell}(\beta, \Lambda_0)(\mathbf{y}) &= (n_0 \mathbf{z} + Q_{\beta, \Lambda_0}(\mathbf{y}) \mathbf{z} \Lambda_0(x)) - \left(n_0 h_{\beta, \Lambda_0}(x) + Q_{\beta, \Lambda_0}(\mathbf{y}) \int_{[\sigma, x]} h_{\beta, \Lambda_0}(s) d\Lambda_0(s) \right) \\ &= n_0 [\mathbf{z} - h_{\beta, \Lambda_0}(x)] + Q_{\beta, \Lambda_0}(\mathbf{y}) \left[\mathbf{z} \Lambda_0(x) - \int_{[\sigma, x]} h_{\beta, \Lambda_0}(s) d\Lambda_0(s) \right] \\ &= n_0 [\mathbf{z} - h_{\beta, \Lambda_0}(x)] + Q_{\beta, \Lambda_0}(\mathbf{y}) \left[\int_{[\sigma, x]} (\mathbf{z} - h_{\beta, \Lambda_0}(s)) d\Lambda_0(s) \right]. \end{aligned} \quad (3.4.21)$$

Ahora encontraremos un submodelo menos favorable tal que cumpla las condiciones (1.3.31) y (1.3.32). Se define el submodelo menos favorable

$$\Lambda_{\mathbf{t}}(x) = \int_{[\sigma, x]} (1 + (\boldsymbol{\beta} - \mathbf{t})' h(s)) d\Lambda_0(s) \quad (3.4.22)$$

para $\Lambda_0(\cdot)$ dada y $h : [\sigma, \tau] \rightarrow \mathbb{R}^d$ con $h(\cdot)$ una dirección menos favorable en los verdaderos valores $(\boldsymbol{\beta}_0, \Lambda_0^0)$.

Nótese que si se toma el submodelo $\Lambda_{\boldsymbol{\beta}_0}(\cdot)$ en $(\boldsymbol{\beta}_0, \Lambda_0^0)$ se tiene

$$\begin{aligned} \Lambda_{\boldsymbol{\beta}_0}(x) &= \int_{[\sigma, x]} (1 + (\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0)' h(s)) d\Lambda_0^0(s) \\ &= \int_{[\sigma, x]} d\Lambda_0^0(s) \\ &= \Lambda_0^0(x) \end{aligned} \quad (3.4.23)$$

esto es, se cumple la condición (1.3.31).

Además, la función de log-verosimilitud del submodelo (3.4.22) esta dada por

$$\begin{aligned} \ell(\mathbf{t}, \boldsymbol{\beta}, \Lambda_0)(\mathbf{y}) &= \log p_{\mathbf{t}, \Lambda_{\mathbf{t}}}(\mathbf{y}) \\ &= n_0 [\mathbf{t}' \mathbf{z} + \log [(1 + (\boldsymbol{\beta} - \mathbf{t})' h(x)) \lambda_0(x)]] - n_{-2} (-\exp(\mathbf{t}' \mathbf{z}) \Lambda_{\mathbf{t}}(x)) \\ &\quad + n_2 \log [1 - \exp(-\Lambda_{\mathbf{t}}(x) \exp(\mathbf{t}' \mathbf{z}))] + \text{cte.} \end{aligned}$$

Así, la función “score” del submodelo es

$$\begin{aligned} \dot{\ell}(\mathbf{t}, \boldsymbol{\beta}, \Lambda_0)(\mathbf{y}) &= \frac{\partial}{\partial \mathbf{t}} \ell(\mathbf{t}, \boldsymbol{\beta}, \Lambda_0) \\ &= n_0 \left[\mathbf{z} + \frac{-h(x) \lambda_0(x)}{(1 + (\boldsymbol{\beta} - \mathbf{t})' h(x)) \lambda_0(x)} \right] \\ &\quad - n_{-2} \left[\exp(\mathbf{t}' \mathbf{z}) \frac{\partial}{\partial \mathbf{t}} \Lambda_{\mathbf{t}}(x) + \Lambda_{\mathbf{t}}(x) \exp(\mathbf{t}' \mathbf{z}) \mathbf{z} \right] \\ &\quad + n_2 \left[\frac{\exp(-\Lambda_{\mathbf{t}}(x) \exp(\mathbf{t}' \mathbf{z})) (\Lambda_{\mathbf{t}}(x) \exp(\mathbf{t}' \mathbf{z}) \mathbf{z} + \exp(\mathbf{t}' \mathbf{z}) \frac{\partial}{\partial \mathbf{t}} \Lambda_{\mathbf{t}}(x))}{1 - \exp(-\Lambda_{\mathbf{t}}(x) \exp(\mathbf{t}' \mathbf{z}))} \right] \end{aligned} \quad (3.4.24)$$

$$\text{con } \frac{\partial}{\partial \mathbf{t}} \Lambda_{\mathbf{t}}(x) = - \int_{[\sigma, x]} h(s) d\Lambda_0(s).$$

Si evaluamos la función anterior en $(\beta_0, \beta_0, \Lambda_0^0)$ se tiene

$$\begin{aligned}
 \dot{\ell}(\beta_0, \beta_0, \Lambda_0^0) &= n_0[\mathbf{z} - h(x)] - n_{-2} \left[\exp(\beta_0' \mathbf{z}) \left(- \int_{[\sigma, x]} h(s) d\Lambda_0^0(s) + \Lambda_0^0(x) \mathbf{z} \right) \right] \\
 &\quad + n_2 \left[\frac{\exp(-\Lambda_0^0(x) \exp(\beta_0' \mathbf{z})) \exp(\beta_0' \mathbf{z})}{1 - \exp(-\Lambda_0^0(x) \exp(\beta_0' \mathbf{z}))} \left(\Lambda_0^0(x) \mathbf{z} - \int_{[\sigma, x]} h(s) d\Lambda_0^0(s) \right) \right] \\
 &= n_0[\mathbf{z} - h(x)] \\
 &\quad + \left(\frac{n_2}{\exp(\Lambda_0^0(x) \exp(\beta_0' \mathbf{z})) - 1} - n_{-2} \right) \exp(\beta_0' \mathbf{z}) \left[\Lambda_0^0(x) \mathbf{z} - \int_{[\sigma, x]} h(s) d\Lambda_0^0(s) \right] \\
 &= n_0[\mathbf{z} - h(x)] + Q_{\beta_0, \Lambda_0^0}(\mathbf{y}) \left[\Lambda_0^0(x) \mathbf{z} - \int_{[\sigma, x]} h(s) d\Lambda_0^0(s) \right]
 \end{aligned}$$

Nótese que $h_{\beta_0, \Lambda_0^0}(\mathbf{x})$ es una dirección menos favorable del modelo, la cual denotaremos por $h_0(\mathbf{x})$, por tanto

$$\begin{aligned}
 \dot{\ell}(\beta_0, \beta_0, \Lambda_0^0)(\mathbf{y}) &= n_0[\mathbf{z} - h_0(x)] + Q_{\beta_0, \Lambda_0^0}(\mathbf{y}) \left[\Lambda_0^0(x) \mathbf{z} - \int_{[\sigma, x]} h_0(s) d\Lambda_0^0(s) \right] \\
 &= n_0[\mathbf{z} - h_0(x)] + Q_{\beta_0, \Lambda_0^0}(\mathbf{y}) \left[\int_{[\sigma, x]} (\mathbf{z} - h_0(s)) d\Lambda_0^0(s) \right],
 \end{aligned}$$

usando (3.4.21) se tiene

$$\dot{\ell}(\beta_0, \beta_0, \Lambda_0^0)(\mathbf{y}) = \tilde{\ell}(\beta_0, \Lambda_0^0)(\mathbf{y}), \quad (3.4.25)$$

esto es, se cumple la condición (1.3.32).

Para verificar la condición (1.3.33) se requiere un estimador de la función de riesgo base acumulativa $\Lambda_0(\cdot)$. Por la construcción del estimador (3.3.13) se asume que la función de riesgo base acumulativa para β dado, se puede estimar por

$$\hat{\Lambda}_\beta(x) = \int_{[0, x]} \frac{dN_0(s)}{E_P [\mathbb{1}_{\{X \geq s, A=0\}} \exp(\beta' \mathbf{z})] + \hat{p} E_P [\mathbb{1}_{\{X \geq s, A=1\}} \exp(\beta' \mathbf{z})]}, \quad (3.4.26)$$

con

$$\hat{p} = \frac{E_P [\mathbb{1}_{\{A=0\}}]}{E_P [\mathbb{1}_{\{A=0\}}] + E_P [\mathbb{1}_{\{A=2\}}]}. \quad (3.4.27)$$

Nótese que el estimador anterior, además de depender del parámetro β , depende de la función de distribución $P(\cdot)$. En el caso de que la función de distribución P

se sustituye por la función de distribución empírica \mathbb{P}_n , el estimador (3.4.26) es el estimador dado en la expresión (3.3.13). Por otra parte, si en el estimador (3.4.27) se toma la verdadera función de distribución $P_0(\cdot)$, se obtiene

$$\begin{aligned}\hat{p} &= \frac{E_0 [\mathbb{1}_{A=0}]}{E_0 [\mathbb{1}_{A=0} + \mathbb{1}_{A=2}]} \\ &= \frac{\int_{\mathbb{R}^d} \int_0^\tau H_0(dt|\mathbf{z})p(\mathbf{z})d\mathbf{z}}{\int_{\mathbb{R}^d} \int_0^\tau (H_0(dt|\mathbf{z}) + H_2(dt|\mathbf{z})) p(\mathbf{z})d\mathbf{z}} \\ &= \frac{\int_{\mathbb{R}^d} H_0([0, \tau]|\mathbf{z})p(\mathbf{z})d\mathbf{z}}{\int_{\mathbb{R}^d} (H_0([0, \tau]|\mathbf{z}) + H_2([0, \tau]|\mathbf{z})) p(\mathbf{z})d\mathbf{z}},\end{aligned}$$

y por la expresión (3.3.7)

$$\begin{aligned}\hat{p} &= \frac{\int_{\mathbb{R}^d} p(H_0([0, \tau]|\mathbf{z}) + H_2([0, \tau]|\mathbf{z}))p(\mathbf{z})d\mathbf{z}}{\int_{\mathbb{R}^d} (H_0([0, \tau]|\mathbf{z}) + H_2([0, \tau]|\mathbf{z}))p(\mathbf{z})d\mathbf{z}} \\ &= p\end{aligned}\tag{3.4.28}$$

Además, si en el estimador (3.4.26) se toma la verdadera función de distribución $P_0(\cdot)$, el verdadero valor del parámetro β_0 y la expresión (3.4.28) se tiene

$$\begin{aligned}\hat{\Lambda}_{\beta_0}(x) &= P_0 \left[\frac{N_0(x)}{E_0 [(\mathbb{1}_{\{X \geq x, A=0\}} + p\mathbb{1}_{\{X \geq x, A=1\}}) \exp(\beta'_0 \mathbf{z})]} \right] \\ &= \int_{\mathbb{R}^d} \int_0^\tau \frac{H_0(dt|\mathbf{z})p(\mathbf{z})d\mathbf{z}}{E_0 [(\mathbb{1}_{\{X \geq x, A=0\}} + p\mathbb{1}_{\{X \geq x, A=1\}}) \exp(\beta'_0 \mathbf{z})]},\end{aligned}$$

y por la expresión (3.3.9) se tiene

$$\begin{aligned}
 \hat{\Lambda}_{\beta_0}(x) &= \int_{\mathbb{R}^d} \int_0^\tau \frac{\exp(\beta'_0 \mathbf{z}) \lambda_0^0(t) dt (H_0([t, \tau]|\mathbf{z}) + p H_1([t, \tau]|\mathbf{z})) p(\mathbf{z}) d\mathbf{z}}{E_0 \left[(\mathbb{1}_{\{X \geq x, A=0\}} + p \mathbb{1}_{\{X \geq x, A=1\}}) \exp(\beta'_0 \mathbf{z}) \right]} \\
 &= \int_{[0, x]} \frac{\lambda_0^0(t) E_0 \left[(\mathbb{1}_{\{X \geq t, A=0\}} + p \mathbb{1}_{\{X \geq t, A=1\}}) \exp(\beta'_0 \mathbf{z}) \right] dt}{E_0 \left[(\mathbb{1}_{\{X \geq x, A=0\}} + p \mathbb{1}_{\{X \geq x, A=1\}}) \exp(\beta'_0 \mathbf{z}) \right]} \\
 &= \int_{[0, x]} \lambda_0^0(t) dt \\
 &= \Lambda_0^0(x).
 \end{aligned}$$

Nótese que el estimador anterior esta definido para $x \in [0, \tau]$ a pesar de que el tiempo observado esta en $[\sigma, \tau]$ con $\sigma > 0$. Ahora se demostrará que el estimador (3.4.26) es consistente, para ello se sigue la metodología de Murphy (1994) usando los siguientes lemas.

Lema 3.4.2. *Bajo las Condiciones 3.4.1, se tiene que*

$$\limsup \hat{\Lambda}_{\beta}(\tau) < \infty$$

casi seguramente.

Demostración. Como el vector \mathbf{z} es acotado, existe una constante c , con $0 < c < \infty$ tal que

$$\|\mathbf{z}\| < c.$$

Para β dado, se define la constante

$$c_2 = \min_{\|\mathbf{z}\| < c} \exp(\beta' \mathbf{z}),$$

entonces

$$\mathbb{P}_n \left[(\mathbb{1}_{\{X \geq x, A=0\}} + \hat{p} \mathbb{1}_{\{X \geq x, A=1\}}) \exp(\beta' \mathbf{z}) \right] \geq c_2 \mathbb{P}_n \left[\mathbb{1}_{\{X \geq x, A=0\}} + \hat{p} \mathbb{1}_{\{X \geq x, A=1\}} \right]$$

y por la ley de los grandes números, se tiene que, casi seguramente $\hat{p} \rightarrow p$

$$\mathbb{P}_n \left[\left(\mathbb{1}_{\{X \geq x, A=0\}} + \hat{p} \mathbb{1}_{\{X \geq x, A=1\}} \right) \exp(\beta' \mathbf{z}) \right] \geq c_2 P_0 \left[\mathbb{1}_{\{X \geq x, A=0\}} + p \mathbb{1}_{\{X \geq x, A=1\}} \right] + o_{P_0}(1).$$

Se puede asumir que, para cada $x \in [\sigma, \tau]$, se tiene

$$P_0 \left[\mathbb{1}_{\{X \geq x, A=0\}} + p \mathbb{1}_{\{X \geq x, A=1\}} \right] > 0,$$

entonces, si $n \rightarrow \infty$

$$\mathbb{P}_n \left[\left(\mathbb{1}_{\{X \geq x, A=0\}} + \hat{p} \mathbb{1}_{\{X \geq x, A=1\}} \right) \exp(\beta' \mathbf{z}) \right] > 0$$

casi seguramente. Entonces como

$$\hat{\Lambda}_\beta(\tau) = \mathbb{P}_n \left[\frac{N_0(x)}{\mathbb{P}_n \left[\left(\mathbb{1}_{\{X \geq x, A=0\}} + \hat{p} \mathbb{1}_{\{X \geq x, A=1\}} \right) \exp(\beta' \mathbf{z}) \right]} \right]$$

se tiene que

$$0 \leq \hat{\Lambda}_\beta(\tau) \leq O(1) \mathbb{P}_n[N_0(\tau)]/c_2$$

casi seguro si $n \rightarrow \infty$, y como $O(1) \mathbb{P}_n[N_0(\tau)]/c_2$ es finito, se obtiene el resultado del lema. □

Lema 3.4.3. *El estimador $\hat{\Lambda}_{\beta_0}(\cdot)$ es consistente, esto es*

$$\sup_{x \in [\sigma, \tau]} \left| \hat{\Lambda}_{\beta_0}(x) - \Lambda_0^0(x) \right| \xrightarrow{c.s.} 0$$

cuando $n \rightarrow \infty$.

Demostración. Sea P_0 la verdadera función de distribución.

Como el vector \mathbf{z} es acotado se tiene que las clases

$$\{\mathbf{z}\} \quad \text{y} \quad \{\beta' \mathbf{z} : \beta \in \Theta\}$$

son Donsker, ya que Θ es un conjunto acotado. Además, por la diferenciabilidad de $\exp(\beta' \mathbf{z})$ y que dichas derivadas son acotadas, se tiene que la clase

$$\{\exp(\beta' \mathbf{z}) : \beta \in \Theta\}$$

es Donsker. También, como las clases

$$\{\mathbb{1}_{\{X \geq x, A=i\}} : x \in [\sigma, \tau]\}$$

con $i = 0, 1, 2$ son Donsker, se tiene que la clase

$$\{\mathbb{1}_{\{X \geq x, A=0\}} + \hat{p}\mathbb{1}_{\{X \geq x, A=1\}} : x \in [\sigma, \tau]\}$$

con $\hat{p} > 0$ es Donsker. Como el producto de dos clases Donsker es Donsker, se obtiene que

$$\{(\mathbb{1}_{\{X \geq x, A=0\}} + \hat{p}\mathbb{1}_{\{X \geq x, A=1\}}) \exp(\beta' \mathbf{z}) : x \in [\sigma, \tau], \beta \in \Theta\} \quad (3.4.29)$$

es Donsker. Debido a que la función $P_0 \rightarrow E_{P_0}[f]$ es Lipschitz la clase

$$\{E_{P_0}[(\mathbb{1}_{\{X \geq x, A=0\}} + \hat{p}\mathbb{1}_{\{X \geq x, A=1\}}) \exp(\beta' \mathbf{z})] : x \in [\sigma, \tau], \beta \in \Theta\}$$

es Donsker. Además, como

$$E_{P_0}[(\mathbb{1}_{\{X \geq x, A=0\}} + \hat{p}\mathbb{1}_{\{X \geq x, A=1\}}) \exp(\beta' \mathbf{z})] > 0$$

y la clase

$$\{N_0(x) = \mathbb{1}_{\{X \leq x, A=0\}} : x \in [\sigma, \tau]\}$$

es Donsker, entonces la clase

$$\left\{ \frac{N_0(x)}{E_{P_0}[(\mathbb{1}_{\{X \geq x, A=0\}} + \hat{p}\mathbb{1}_{\{X \geq x, A=1\}}) \exp(\beta' \mathbf{z})]} : x \in [\sigma, \tau], \beta \in \Theta \right\} \quad (3.4.30)$$

es Donsker.

Por otra parte, si se define $M_q(x) = \mathbb{1}_{\{X \geq x, A=0\}} + q\mathbb{1}_{\{X \geq x, A=1\}}$, se tiene

$$\begin{aligned} \sup_{x \in [\sigma, \tau]} |\hat{\Lambda}_{\beta_0}(x) - \Lambda_0^0(x)| &= \sup_{x \in [\sigma, \tau]} \left| \mathbb{P}_n \left[\frac{N_0(x)}{\mathbb{P}_n[M_{\hat{p}}(x) \exp(\beta_0' \mathbf{z})]} \right] - P_0 \left[\frac{N_0(x)}{P_0[M_p(x) \exp(\beta_0' \mathbf{z})]} \right] \right| \\ &\leq \sup_{x \in [\sigma, \tau]} \left| \mathbb{P}_n \left[\frac{N_0(x)}{\mathbb{P}_n[M_{\hat{p}}(x) \exp(\beta_0' \mathbf{z})]} \right] - \mathbb{P}_n \left[\frac{N_0(x)}{P_0[M_p(x) \exp(\beta_0' \mathbf{z})]} \right] \right| \\ &\quad + \sup_{x \in [\sigma, \tau]} \left| \mathbb{P}_n \left[\frac{N_0(x)}{P_0[M_p(x) \exp(\beta_0' \mathbf{z})]} \right] - P_0 \left[\frac{N_0(x)}{P_0[M_p(x) \exp(\beta_0' \mathbf{z})]} \right] \right| \end{aligned}$$

tomando $M_q^i(x) = \mathbb{1}_{\{X_i \geq x, A=0\}} + q\mathbb{1}_{\{X_i \geq x, A=1\}}$ y $N_0^i(x) = \mathbb{1}_{\{X_i \leq x, A=0\}}$ se obtiene

$$\begin{aligned} \sup_{x \in [\sigma, \tau]} \left| \hat{\Lambda}_{\beta_0}(x) - \Lambda_0^0(x) \right| &\leq \sup_{x \in [\sigma, \tau]} \left| \frac{1}{n} \sum_{i=1}^n N_0^i(x) \left[\frac{1}{\mathbb{P}_n \left[M_{\hat{p}}^i(x) \exp(\beta'_0 \mathbf{z}) \right]} - \frac{1}{P_0 \left[M_p^i(x) \exp(\beta'_0 \mathbf{z}) \right]} \right] \right| \\ &\quad + \sup_{x \in [\sigma, \tau]} \left| (\mathbb{P}_n - P_0) \left[\frac{N_0(x)}{P_0 \left[M_p(x) \exp(\beta'_0 \mathbf{z}) \right]} \right] \right| \\ &\leq \sup_{x \in [\sigma, \tau]} \left| \frac{1}{\mathbb{P}_n \left[M_{\hat{p}}^i(x) \exp(\beta'_0 \mathbf{z}) \right]} - \frac{1}{P_0 \left[M_p^i(x) \exp(\beta'_0 \mathbf{z}) \right]} \right| \\ &\quad + \sup_{x \in [\sigma, \tau]} \left| (\mathbb{P}_n - P_0) \left[\frac{N_0(x)}{P_0 \left[M_p(x) \exp(\beta'_0 \mathbf{z}) \right]} \right] \right| \\ &\leq \sup_{x \in [\sigma, \tau]} \left| \frac{1}{\mathbb{P}_n \left[M_{\hat{p}}^i(x) \exp(\beta'_0 \mathbf{z}) \right]} - \frac{1}{P_0 \left[M_{\hat{p}}^i(x) \exp(\beta'_0 \mathbf{z}) \right]} \right| \\ &\quad + \sup_{x \in [\sigma, \tau]} \left| \frac{1}{P_0 \left[M_{\hat{p}}^i(x) \exp(\beta'_0 \mathbf{z}) \right]} - \frac{1}{P_0 \left[M_p^i(x) \exp(\beta'_0 \mathbf{z}) \right]} \right| \\ &\quad + \sup_{x \in [\sigma, \tau]} \left| (\mathbb{P}_n - P_0) \left[\frac{N_0(x)}{P_0 \left[M_p(x) \exp(\beta'_0 \mathbf{z}) \right]} \right] \right| \end{aligned}$$

Como la clase (3.4.30) es Glivenko–Cantelli, se tiene

$$\sup_{x \in [\sigma, \tau]} \left| (\mathbb{P}_n - P_0) \left[\frac{N_0(x)}{P_0 \left[M_p(x) \exp(\beta'_0 \mathbf{z}) \right]} \right] \right| \xrightarrow{c.s.} 0.$$

Además, por la ley de grandes números y el Teorema central del límite se tiene

$$\hat{p} \xrightarrow{c.s.} p,$$

entonces, como $P_0 \left[M_p^i(x) \exp(\beta'_0 \mathbf{z}) \right] > 0$

$$\sup_{x \in [\sigma, \tau]} \left| \frac{1}{P_0 \left[M_{\hat{p}}^i(x) \exp(\beta'_0 \mathbf{z}) \right]} - \frac{1}{P_0 \left[M_p^i(x) \exp(\beta'_0 \mathbf{z}) \right]} \right| \xrightarrow{c.s.} 0.$$

Finalmente, como la clase (3.4.29) es Glivenko–Cantelli se tiene

$$\sup_{x \in [\sigma, \tau]} \left| (\mathbb{P}_n - P_0) \left[M_{\hat{p}}^i(x) \exp(\beta'_0 \mathbf{z}) \right] \right| \xrightarrow{c.s.} 0$$

y como $P_0 \left[M_{\hat{p}}^i(x) \exp(\beta'_0 \mathbf{z}) \right] > 0$ para $x \in [\sigma, \tau]$

$$\sup_{x \in [\sigma, \tau]} \left| \frac{1}{\mathbb{P}_n \left[M_{\hat{p}}^i(x) \exp(\beta'_0 \mathbf{z}) \right]} - \frac{1}{P_0 \left[M_{\hat{p}}^i(x) \exp(\beta'_0 \mathbf{z}) \right]} \right| \xrightarrow{c.s.} 0.$$

Por tanto, se tiene la demostración del lema. \square

Teorema 3.4.4. *Bajo las Condiciones 3.4.1, el estimador $\tilde{\beta}_n$ de máxima log-verosimilitud “profile” del modelo de Cox (3.3.3) con datos doblemente censurados es consistente, es decir*

$$\|\tilde{\beta}_n - \beta_0\| \xrightarrow{c.s.} 0.$$

Además, el estimador de la función de riesgo base $\hat{\Lambda}_{\tilde{\beta}_n}(\cdot)$ es consistente esto es

$$\sup_{x[\sigma, \tau]} \left| \hat{\Lambda}_{\tilde{\beta}_n}(x) - \Lambda_0^0(x) \right| \xrightarrow{c.s.} 0.$$

Demostración. La demostración es análoga a la del Teorema 2.6.6. □

Como consecuencia del teorema anterior se cumple la Condición (1.3.33). Finalmente, en el siguiente lema se verifica que la Condición (1.3.34) se cumple.

Lema 3.4.5. *Bajo la consistencia de los estimadores $\tilde{\beta}_n$ y $\hat{\Lambda}_{\tilde{\beta}_n}(\cdot)$ se cumple*

$$P_0 \left[\dot{\ell}(\beta_0, \tilde{\beta}_n, \hat{\Lambda}_{\tilde{\beta}_n}) \right] = o_P \left(\|\tilde{\beta}_n - \beta_0\| + n^{-\frac{1}{2}} \right) \quad (3.4.31)$$

Demostración. Bajo las Condiciones 3.4.1, de forma análoga a Murphy y van der Vaart (1997) se puede ver que

$$\|\hat{\Lambda}_{\tilde{\beta}_n} - \Lambda_0^0\|_2 = O_P(n^{-2/3}) \quad (3.4.32)$$

donde $\|\cdot\|_2$ es la L_2 -norma en $[\sigma, \tau]$.

Considerando el modelo (3.4.22) y la dirección menos favorable

$$h_0(x) = \frac{M_1(x)}{M_0(x)}$$

con

$$\begin{aligned} M_0(x) &= 2\Lambda_0(x) \left[E \left[(Q_{\beta, \Lambda_0}(\mathbf{y}))^2 \mathbb{1}_{\{X \geq x\}} \right] - E \left[\exp(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}} \right] \right] \\ &\quad + E \left[\exp(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}} \right], \end{aligned} \quad (3.4.33)$$

$$\begin{aligned} M_1(x) &= 2\Lambda_0(x) \left[E \left[\mathbf{z} (Q_{\beta, \Lambda_0}(\mathbf{y}))^2 \mathbb{1}_{\{X \geq x\}} \right] - E \left[\mathbf{z} \exp(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}} \right] \right] \\ &\quad + E \left[\mathbf{z} \exp(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}} \right], \end{aligned} \quad (3.4.34)$$

nótese que $M_0(\cdot)$ y $M_1(\cdot)$ dependen de los parámetros de β y Λ_0 . Se tiene que la log-verosimilitud del dato \mathbf{y} viene dada por

$$\ell(\mathbf{t}, \beta, \Lambda_0)(\mathbf{y}) = \log p_{\mathbf{t}, \Lambda_0}(\mathbf{y}).$$

Usando la expresión (3.4.24) se observa que la función “score” del submodelo con respecto a \mathbf{t} está dada por

$$\begin{aligned} \dot{\ell}(t, \beta, \Lambda_0)(\mathbf{y}) &= n_0 \left[\mathbf{z} - \frac{h_0(x)}{1 + (\beta - \mathbf{t})' h_0(x)} \right] \\ &\quad + Q(\mathbf{y}) \left[\mathbf{z} \int_{\sigma}^x (1 + (\beta - \mathbf{t})' h_0(s)) d\Lambda_0(s) - \int_{\sigma}^x h_0(s) d\Lambda_0(s) \right] \end{aligned}$$

$$\text{con } Q(\mathbf{y}) = Q_{\mathbf{t}, \beta, \Lambda_0}(\mathbf{y}) = \left(\frac{n_2}{\exp(\Lambda_{\mathbf{t}}(x) \exp(\mathbf{t}' \mathbf{z})) - 1} - n_{-2} \right) \exp(\mathbf{t}' \mathbf{z}).$$

Entonces,

$$\begin{aligned} \dot{\ell}(\beta_0, \beta_0, \Lambda_0) &= n_0 [\mathbf{z} - h_0(x)] + Q(\mathbf{y}) \left[\mathbf{z} \int_{\sigma}^x d\Lambda_0(s) - \int_{\sigma}^x h_0(s) d\Lambda_0(s) \right] \\ &= n_0 [\mathbf{z} - h_0(x)] + Q(\mathbf{y}) \left[\int_{\sigma}^x (\mathbf{z} - h_0(s)) d\Lambda_0(s) \right], \end{aligned}$$

$$\text{con } Q(\mathbf{y}) = Q_{\beta_0, \Lambda_0}(\mathbf{y}).$$

Para simplificar la notación, se escribirá

$$\dot{\ell}(\Lambda) = \dot{\ell}(\beta_0, \beta_0, \Lambda).$$

Así se tiene

$$P_0[\dot{\ell}(\hat{\Lambda}_{\tilde{\beta}_n})] = (P_0 - P_{\beta_0, \hat{\Lambda}_{\tilde{\beta}_n}}) \dot{\ell}(\Lambda_0^0) + (P_0 - P_{\beta_0, \hat{\Lambda}_{\tilde{\beta}_n}})(\dot{\ell}(\hat{\Lambda}_{\tilde{\beta}_n}) - \dot{\ell}(\Lambda_0^0)), \quad (3.4.35)$$

ya que $P_{\beta_0, \hat{\Lambda}_{\tilde{\beta}_n}} \dot{\ell}(\hat{\Lambda}_{\tilde{\beta}_n}) = 0$.

Como $\dot{\ell}(\Lambda_0^0)$ es la función “score” eficiente para β , es ortogonal para cada función “score” de Λ , entonces

$$(P_0 - P_{\beta_0, \hat{\Lambda}_{\tilde{\beta}_n}}) \dot{\ell}(\Lambda_0^0) = P_0 \dot{\ell}(\Lambda_0^0) \left[\frac{(p_0 - p_{\beta_0, \hat{\Lambda}_{\tilde{\beta}_n}})}{p_0} - \ell(\beta_0, \Lambda_0^0)(\Lambda_0^0 - \hat{\Lambda}_{\tilde{\beta}_n}) \right].$$

Nótese que $\left[\frac{(p_0 - P_{\beta_0, \hat{\Lambda}_{\tilde{\beta}_n}})}{p_0} - \ell(\beta_0, \Lambda_0^0)(\Lambda_0^0 - \hat{\Lambda}_{\tilde{\beta}_n}) \right]$ es una aproximación lineal de $\Lambda_0^0 - \hat{\Lambda}_{\tilde{\beta}_n}$. Considerando el desarrollo de Taylor en un orden más se tiene que este factor es acotado por un múltiplo de $(\Lambda_0^0 - \hat{\Lambda}_{\tilde{\beta}_n})^2$ entonces $(P_0 - P_{\beta_0, \hat{\Lambda}_{\tilde{\beta}_n}}) \dot{\ell}(\Lambda_0^0)$ está acotada por un múltiplo de $P_0(\Lambda_0^0 - \hat{\Lambda}_{\tilde{\beta}_n})^2$, es decir, por (3.4.32) es $O_P(n^{-\frac{2}{3}})$.

Por otro lado, el segundo término de (3.4.35) se puede acotar similarmente porque $\Lambda \mapsto P_{\beta_0, \Lambda}$ y $\Lambda \mapsto \dot{\ell}(\beta_0, \beta_0, \Lambda)$ son funciones uniformemente Lipschitz.

Esto sucede cuando el submodelo es evaluado en $(\beta_0, \hat{\Lambda}_{\tilde{\beta}_n})$ en lugar de $(\tilde{\beta}_n, \hat{\Lambda}_{\tilde{\beta}_n})$, pero tomando la diferencia de estas dos expresiones es $o_P(\|\tilde{\beta}_n - \beta_0\|)$. Por tanto se cumple el lema. \square

Ahora, se verifican las hipótesis del Teorema 1.3.10 las cuales están escritas en el siguiente lema.

Lema 3.4.6. *Sea V una vecindad de $(\beta_0, \beta_0, \Lambda_0)$ donde β_0 y $\Lambda_0^0(\cdot)$ son los verdaderos parámetros del modelo de Cox (3.3.3) con datos doblemente censurados y $\ell(\mathbf{t}, \beta, \Lambda_0)$ es la función de log-verosimilitud del submodelo (3.4.22).*

i) *La clase de funciones*

$$\{\dot{\ell}(\mathbf{t}, \beta, \Lambda_0) : (\mathbf{t}, \beta, \Lambda_0) \in V\}$$

es Donsker con función envolvente cuadrado integrable, y

ii) *la clase de funciones*

$$\{\ddot{\ell}(\mathbf{t}, \beta, \Lambda_0) : (\mathbf{t}, \beta, \Lambda_0) \in V\}$$

es Glivenko–Cantelli con función envolvente integrable.

Demostración. Sea \mathcal{F} el conjunto de funciones de distribuciones continuas. Para $\rho > 0$ se define

$$C_\rho = \{P \in \mathcal{F} : \|P - P_0\|_\infty \leq \rho\}$$

con P_0 la verdadera función de distribución.

Como se sabe, las clases

$$\{\mathbf{z}\}, \quad \{\exp(\boldsymbol{\beta}'\mathbf{z}) : \boldsymbol{\beta} \in \Theta\} \quad \text{y} \quad \{\exp^2(\boldsymbol{\beta}'\mathbf{z}) : \boldsymbol{\beta} \in \Theta\}$$

son Donsker debido a que el vector \mathbf{z} es acotado, Θ es un compacto y la diferenciabilidad de $\exp(\boldsymbol{\beta}'\mathbf{z})$, además de que dichas derivadas son acotadas.

Además como $\Lambda_0(\cdot)$ es uniformemente acotada y de variación uniformemente acotada

$$\{\Lambda_0(x) : x \in [\sigma, \tau]\}$$

es Donsker, lo cual implica que la clase

$$\{\Lambda_0(x) \exp(\boldsymbol{\beta}'\mathbf{z}) : x \in [\sigma, \tau], \boldsymbol{\beta} \in \Theta\}$$

es Donsker. Por la diferenciabilidad de la función exponencial, y a que dichas derivadas son acotadas se tiene que

$$\{\exp(\Lambda_0(x) \exp(\boldsymbol{\beta}'\mathbf{z})) : x \in [\sigma, \tau], \boldsymbol{\beta} \in \Theta\}$$

es Donsker. Dado que $\Lambda_0(\cdot) > 0$ en $[\sigma, \tau]$, si $\exp(\boldsymbol{\beta}'\mathbf{z})$ para $\boldsymbol{\beta} \in \Theta$ se tiene que $\exp(\Lambda_0(x) \exp(\boldsymbol{\beta}'\mathbf{z})) - 1 > 0$, entonces la clase

$$\left\{ \frac{1}{\exp(\Lambda_0(x) \exp(\boldsymbol{\beta}'\mathbf{z})) - 1} : x \in [\sigma, \tau], \boldsymbol{\beta} \in \Theta \right\}$$

es Donsker, aplicando las propiedades de permanencia de clase Donsker se tiene que

$$\left\{ Q_{\boldsymbol{\beta}, \Lambda_0}(\mathbf{y}) = \left(\frac{n_2}{\exp(\Lambda_0(x) \exp(\boldsymbol{\beta}'\mathbf{z})) - 1} - n_{-2} \right) \exp(\boldsymbol{\beta}'\mathbf{z}) : x \in [\sigma, \tau], \boldsymbol{\beta} \in \Theta \right\} \quad (3.4.36)$$

es Donsker, entonces la clase

$$\{Q_{\boldsymbol{\beta}, \Lambda_0}^2(\mathbf{y}) : x \in [\sigma, \tau], \boldsymbol{\beta} \in \Theta\}$$

también lo es. Como la clase

$$\{\mathbb{1}_{\{X \geq x\}} : x \in [\sigma, \tau]\}$$

es Donsker y para $P \in C_\rho$, la función $P \rightarrow E_P[f]$, es Lipschitz, las clases

$$\{M_1(x) : x \in [\sigma, \tau], \beta \in \Theta, P \in C_\rho\}$$

y

$$\{M_0(x) : x \in [\sigma, \tau], \beta \in \Theta, P \in C_\rho\}$$

son Donsker, donde $M_0(\cdot)$ y $M_1(\cdot)$ están definidas en (3.4.33) y (3.4.34), respectivamente.

Nótese que

$$Q_{\beta, \Lambda_0}^2(y) = \left(\frac{n_2}{(\exp(\Lambda_0(x) \exp(\beta' \mathbf{z})) - 1)^2} + n_{-2} \right) \exp^2(\beta' \mathbf{z})$$

entonces, si $n_{-2} = 1$

$$\exp^2(\beta' \mathbf{z}) = Q_{\beta, \Lambda_0}^2(\mathbf{y})$$

y si $n_2 = 1$

$$\exp^2(\beta' \mathbf{z}) \leq Q_{\beta, \Lambda_0}^2(\mathbf{y})$$

ya que $\exp(\Lambda_0(x) \exp(\beta' \mathbf{z}))$ es la función de supervivencia y toma valores en $[0, 1)$.

De donde

$$\begin{aligned} 0 &< \exp^2(\beta' \mathbf{z}) \leq Q_{\beta, \Lambda_0}^2(\mathbf{y}), \\ 0 &< \exp^2(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}} \leq Q_{\beta, \Lambda_0}^2(\mathbf{y}) \mathbb{1}_{\{X \geq x\}}, \\ 0 &< E [\exp^2(\beta' \mathbf{z}) \mathbb{1}_{\{X \geq x\}}] \leq E [Q_{\beta, \Lambda_0}^2(\mathbf{y}) \mathbb{1}_{\{X \geq x\}}], \end{aligned}$$

por tanto

$$0 < M_0(x)$$

y entonces

$$\left\{ h_0(x) = \frac{M_1(x)}{M_0(x)} : \beta \in \Theta, x \in [\sigma, \tau], P \in C_\rho \right\}$$

es Donsker. Por consiguiente, la clase

$$\{\mathbf{z} - h_0(x) : \boldsymbol{\beta} \in \Theta, x \in [\sigma, \tau], P \in C_\rho\} \quad (3.4.37)$$

es Donsker. Como la función $\Lambda_0 \rightarrow \int f d\Lambda_0$ es Lipschitz, la clase

$$\left\{ \int_{[\sigma, x]} (\mathbf{z} - h_0(s)) d\Lambda_0(s) : x \in [\sigma, \tau], \boldsymbol{\beta} \in \Theta, P \in C_\rho \right\} \quad (3.4.38)$$

es Donsker.

Finalmente, como consecuencia de (3.4.36), (3.4.37) y (3.4.38) se obtiene que

$$\left\{ \tilde{\ell}(\boldsymbol{\beta}, \Lambda_0) : x \in [\sigma, \tau], \boldsymbol{\beta} \in \Theta, P \in C_\rho \right\}$$

es Donsker, y por tanto, con función envolvente cuadrado integrable, con la cual se obtiene **i)**. Por otra parte, la demostración de **ii)** es análoga. \square

Por último, se puede demostrar el siguiente teorema de normalidad asintótica del estimador del modelo.

Teorema 3.4.7. *Bajo las Condiciones 3.4.1 el estimador $\tilde{\boldsymbol{\beta}}_n$ de máxima verosimilitud “profile” del modelo de Cox (3.3.3) con datos doblemente censurados se distribuye*

$$\sqrt{n}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(\mathbf{0}, \tilde{\mathbf{I}}_0^{-1}),$$

donde $\boldsymbol{\beta}_0$ es el verdadero valor del parámetro del modelo y la matriz $\tilde{\mathbf{I}}_0$ viene dada por $P_0[(\dot{\ell}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0, \Lambda_0^0))(\dot{\ell}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0, \Lambda_0^0))']$ con $\dot{\ell}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0, \Lambda_0^0)$ como en (3.4.25).

Demostración. En las expresiones (3.4.23) y (3.4.25) se verifican las condiciones (1.3.31) y (1.3.32), respectivamente. Por el Teorema 3.4.4 se cumple la condición (1.3.33) de consistencia de los estimadores y por el Lema 3.4.5 se cumple la condición (1.3.34) de no-sesgo del estimador. Además, por el Lema 3.4.6 se puede aplicar el Teorema 1.3.10. Finalmente por la condición vi) de las Condiciones 3.4.1 se tiene el resultado del Corolario 1.3.11, esto es, $\sqrt{n}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ se distribuye asintóticamente como una $\mathcal{N}(\mathbf{0}, \tilde{\mathbf{I}}_0^{-1})$, con $\boldsymbol{\beta}_0$ el verdadero valor del parámetro del modelo y la matriz $\tilde{\mathbf{I}}_0 = P_0[(\dot{\ell}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0, \Lambda_0^0))(\dot{\ell}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0, \Lambda_0^0))']$. \square

3.5. Aplicación a Datos Reales

En esta sección, se utilizan los datos de supervivencia analizados en Lau y otros (2009), estos datos son un fragmento de los datos de “The Women’s Interagency HIV Study” (WIHS). Dicho estudio, se estableció en Agosto de 1993 para estudiar el impacto del virus de inmunodeficiencia humana (VIH) en mujeres estadounidenses de 6 lugares diferentes: Nueva York (2 sitios), Washington D.C., Los Ángeles, San Francisco y Chicago. Entre 1994 y 1995, entraron en el estudio 2,054 mujeres VIH-positivo y 569 mujeres VIH-negativo que fueron examinadas físicamente, entrevistadas y, además se hicieron pruebas de laboratorio en intervalos de 6 meses para obtener los datos de supervivencia sobre ellas.

El fragmento de datos analizado son datos de 1,164 mujeres infectadas con VIH pero sin presencia del síndrome de inmunodeficiencia adquirida (SIDA) el 6 de Diciembre de 1995. En esta fecha, la Agencia de Drogas y Alimentos (Federal Drugs Administration, FDA) aprobó el primer inhibidor de la protasa. La información que se tiene de las mujeres del estudio clínico es la siguiente: raza afroamericana (0=no, 1=sí), edad, historial de uso de drogas (0=no, 1=sí) y la cantidad de la molécula CD4 (por 100 células/uL). El tiempo registrado para cada individuo es el tiempo de diagnóstico de SIDA, tiempo de muerte, tiempo de inicio del tratamiento sin presencia de SIDA o tiempo de fin del estudio clínico (28 de Septiembre de 2006). Los datos se encuentran disponibles en el paquete `randomSurvivalForest` del software libre R con el nombre de `wihs` en cran.r-project.org/web/packages/randomSurvivalForest/index.html.

En este estudio, el evento de interés es del diagnóstico de SIDA o la muerte del individuo. Además, si el tiempo registrado para un individuo es el tiempo de inicio del tratamiento el dato es considerado como un dato con censura por la izquierda y si el tiempo registrado es el tiempo de fin del estudio el dato es considerado como

un dato con censura por la derecha.

Estudios previos muestran que un individuo VIH-positivo con historial de uso de drogas es menos probable que inicie un tratamiento efectivo y más probable que muera en la etapa de terapia con antiretrovirales que los que no tienen historial de uso de drogas. En Lau y otros (2009), se hace un análisis de supervivencia en el sentido antes descrito y por ello los datos son divididos en grupos, esto es, se hacen análisis por separado para el grupo sin y con historial de uso de drogas. En la Tabla 3.1 se resume la información de los datos por grupos. Se puede observar que en el grupo de los individuos con historial con drogas hay mayor proporción de mujeres de raza afroamericana, la edad promedio y el valor promedio de CD4 son mayores. Nótese que en el grupo de individuos con historial sin drogas la proporción de mujeres que mueren o presentan SIDA es menor, además la proporción de mujeres que inician el tratamiento es mayor.

En este trabajo se analizan los datos WIHS en un sentido diferente, como se ha mencionado antes, considerando datos doblemente censurado, sin embargo el análisis de datos también se puede hacer por grupos. En la Tabla 3.2, se muestra la estimación de los parámetros del modelo de Cox, así como su error estándar (EE), tanto para el modelo ajustado a todos los datos, como a los datos con y sin historial de drogas por separado. Se puede observar que el efecto del valor de CD4 es del mismo signo en ambos grupos, en cambio el efecto de la edad y raza son de diferente signo para cada grupo. Además, los signos de los efectos de las covariables para el grupo de mujeres sin historial de drogas coincide con los de la muestra completa.

En la Figura 3.1, se muestra la función de supervivencia estimada: en la figura A) se usan las 4 covariables, mientras que en la figura B) se hace la estimación por grupos (con y sin historial de drogas) y se puede ver que los individuos con historial de uso de drogas tienen menor supervivencia que los que no.

Modelo de Cox para Datos Doblemente Censurados

	Todos n=1164	Historial sin drogas n=725	Historial con drogas n=430
Media de edad (años)	36	34	40
Media de CD4 (100 cel./uL)	3.94	3.87	4.04
Raza (% afroamericanos)	57.73	55.04	62.19
% datos completos	30.84	23.31	43.28
% datos censura derecha	10.82	12.00	8.88
% datos censura izquierda	58.33	64.69	47.84

Tabla 3.1: Información sobre los datos de infección de VIH.

Variable	Todos		Historial c/ drogas		Historial s/ drogas	
	Est	EE	Est	EE	Est	EE
Drogas	-0.122	0.129	—	—	—	—
Edad	0.033	0.009	-0.013	0.013	0.065	0.010
Raza	-0.187	0.116	0.158	0.156	-0.600	0.145
CD4	-0.357	0.023	-0.217	0.030	-0.494	0.034

Tabla 3.2: Estimación del parámetro y error estándar para datos de infección de VIH.

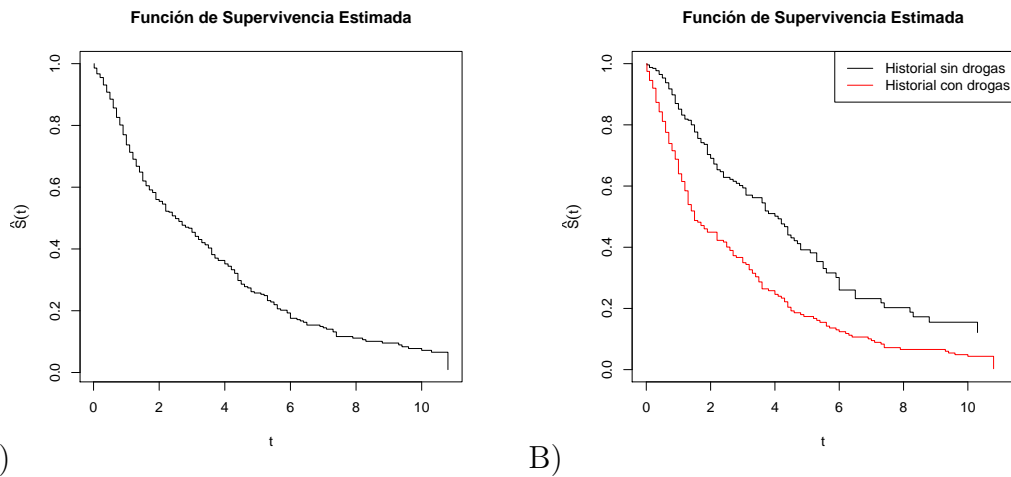


Figura 3.1: Gráfica de la estimación de la función de supervivencia para datos de infección de VIH A) de toda la muestra y B) de la muestra dividida en dos grupos: con y sin historial de drogas.

Por otro lado, supóngase que no se puede ajustar el modelo de Cox a datos doblemente censurados y sólo se puede ajustar a datos con censura por la derecha. En ese caso se tomaría un fragmento de los datos (datos completos y datos con censura por la derecha) dejando de lado la información que aportan los datos con censura por la izquierda. Como se puede ver en la Tabla 3.1 la proposición de datos con censura por la izquierda es cerca de la mitad. Ahora, ajustaremos el modelo de Cox al fragmento de datos completos y con censura por la derecha para observar la pérdida de información. En la Tabla 3.3 se resume la información de los datos. Nótese que en este caso la proporción de datos completos es mucho mayor. En la Tabla 3.4 se muestra la estimación de los parámetros y sus errores estándar (EE) para el modelo de Cox, además también se hace un ajuste por grupos. En la Figura 3.2 A y B se muestra la estimación de la función de supervivencia para 4 covariables y por grupo (con/sin historial de drogas), respectivamente. Al igual que en el análisis anterior se observa que los individuos con historial de drogas tienen mayor riesgo de morir. Sin embargo, si hay un cambio importante, el signo del parámetro de la variable drogas es diferente. Hasta aquí, parece que usar los datos doblemente censurados casi no han aportado mucho al modelo. Sin embargo, si se analiza el riesgo relativo se obtienen diferentes conclusiones. Se analiza el riesgo relativo de un par de individuos con el mismo valor de CD4 y de la misma raza considerando dos casos que tienen o no historial de uso de drogas y cuya diferencia de edad sea 10 años, el riesgo relativo es 0.99 y 1.35 en el caso de sin uso de drogas y con uso de drogas, respectivamente. Análogamente, bajo el modelo de Cox para datos doblemente censurados el riesgo relativo es 0.88 y 1.92, sin y con uso de drogas respectivamente. Se observa, que para el caso de uso de drogas la diferencia en el riesgo relativo es significativa. Es decir, considerando los datos doblemente censurados se concluye que el riesgo de ser diagnosticado de SIDA o de muerte habiendo consumido drogas es mucho mayor que no habiendo consumido

Modelo de Cox para Datos Doblemente Censurados

	Todos <i>n</i> =485	Historial sin drogas <i>n</i> =256	Historial con drogas <i>n</i> = 229
Media de edad (años)	37	34	40
Media de CD4 (100 cel./uL)	4.24	4.28	4.20
Raza (% afroamericanos)	63.71	63.28	64.19
% datos completos	74	66	83
% datos censura derecha	26	34	17

Tabla 3.3: Información sobre los datos de infección de VIH considerando sólo datos con censura por la derecha.

Variable	Todos		Historial c/ drogas		Historial s/ drogas	
	Est	EE	Est	EE	Est	EE
Drogas	0.132	0.118	—	—	—	—
Edad	0.017	0.008	-0.001	0.013	0.030	0.011
Raza	-0.100	0.111	0.011	0.157	-0.208	0.162
CD4	-0.202	0.022	-0.139	0.029	-0.280	0.035

Tabla 3.4: Estimación del parámetro y error estándar para datos con censura por la derecha de infección de VIH.

si se consideran los datos completos y censurados por la derecha, es decir, sólo los datos de aquellos individuos para los que el tiempo observado es el abandono o fin del estudio o la presencia de SIDA o la muerte.

3.6. Estudio de Simulación

Una vez demostrada la normalidad asintótica de los estimadores de los parámetros del modelo de Cox para datos doblemente censurados, se verá el comportamiento de éstos para muestras finitas. En esta sección, se presentan los resultados obtenidos a partir de datos simulados.

El tiempo de supervivencia t se genera bajo el modelo

$$\lambda(t|\mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z}),$$

donde la función de riesgo base $\lambda_0(\cdot)$ es una función Weibull con parámetros $\gamma =$

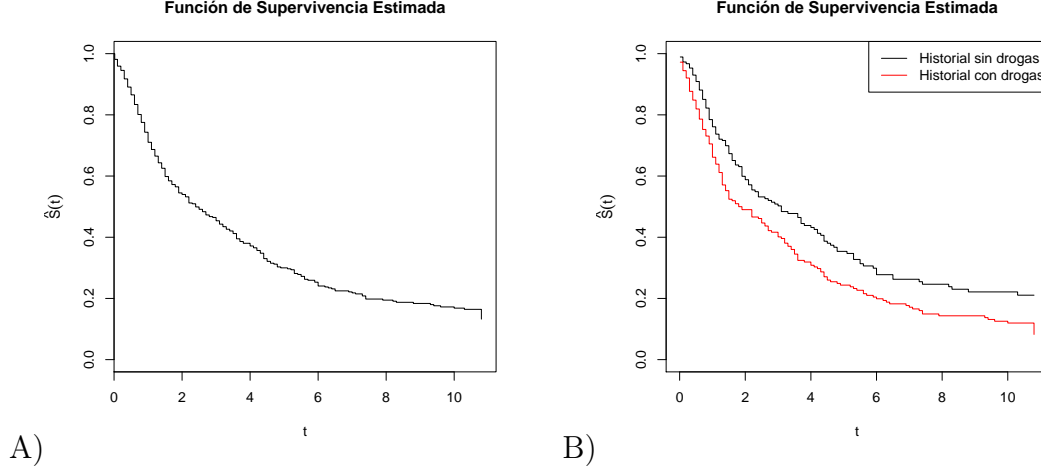


Figura 3.2: Gráfica de la estimación de la función de supervivencia para datos con censura por la derecha de infección de VIH A) de toda la muestra y B) de la muestra dividida en dos grupos: con y sin historial de drogas.

3.2 y $\nu = 1.1$, entonces, por (1.2.25), el tiempo de supervivencia se obtiene usando la expresión

$$t = \left(\frac{-1}{\gamma} \frac{\log u}{\exp(\beta' \mathbf{z})} \right)^{\frac{1}{\nu}}$$

con $u \sim \mathcal{U}(0,1)$. El tiempo de censura c se genera bajo otra distribución y no depende de \mathbf{z} .

Se han diseñado experimentos con diferentes tipos de covariables, primero se realizan experimentos con una covariable y, posteriormente, se generan datos con dos covariables. Para cada diseño de experimento se consideran tamaños de muestra de $n = 50, 100, 400$ y 600 y se realizaron 5,000 réplicas para cada configuración.

Caso 1–dimensional

En este caso, se escoge $\beta_0 = 1$ con $Z \sim \mathcal{U}(0,1)$ y se generan los tiempos T , posteriormente se fija el valor de p y se generan los tiempos C , de tal modo que se obtengan datos de la forma (3.2.1) con diferentes proporciones de datos

Modelo de Cox para Datos Doblemente Censurados

Diseño	n	Media	Mediana	EE	E _E	p-valor	PC para 0.90	PC para 0.95	PC para 0.99
1	50	1.08548	1.05873	0.63099	0.67916	0.00000	0.881	0.936	0.987
	100	1.02968	1.02404	0.42895	0.43963	0.00154	0.894	0.944	0.991
	400	1.02566	1.02331	0.20942	0.21556	0.01061	0.891	0.946	0.986
	600	1.01984	1.01881	0.17036	0.17219	0.25675	0.899	0.945	0.989
2	50	1.08956	1.06118	0.75128	0.80821	0.00000	0.886	0.939	0.983
	100	1.05849	1.03687	0.50608	0.52993	0.00000	0.896	0.942	0.987
	400	1.03611	1.03205	0.24444	0.24918	0.27481	0.896	0.944	0.988
	600	1.03442	1.02981	0.19885	0.20259	0.20369	0.892	0.943	0.988
3	50	1.12689	1.05766	0.94683	1.06670	0.00000	0.877	0.930	0.980
	100	1.07317	1.05383	0.62511	0.66125	0.00000	0.890	0.941	0.986
	400	1.03923	1.03021	0.29654	0.30441	0.00013	0.892	0.944	0.987
	600	1.03807	1.03625	0.24062	0.24574	0.00221	0.895	0.942	0.988
4	50	1.11156	1.07911	0.72339	0.78459	0.00000	0.878	0.936	0.985
	100	1.10427	1.09370	0.49176	0.52419	0.01150	0.879	0.936	0.986
	400	1.07023	1.06746	0.23743	0.24988	0.06432	0.868	0.933	0.983
	600	1.06229	1.05964	0.19278	0.19931	0.04144	0.874	0.932	0.981
5	50	1.18929	1.14129	0.87130	0.99962	0.00000	0.856	0.919	0.976
	100	1.15642	1.14672	0.57858	0.63732	0.00000	0.868	0.926	0.978
	400	1.11057	1.10571	0.27675	0.28651	0.65006	0.863	0.923	0.982
	600	1.10951	1.10796	0.22505	0.23172	0.84528	0.854	0.916	0.979

Tabla 3.5: Resultados para 5000 réplicas con diferentes proporciones de censura y diferentes tamaños de muestra (datos con una sola covariable).

completos, datos con censura por la izquierda y datos con censura por la derecha, respectivamente. Así, en la Tabla 3.5 y Figura 3.3 se muestran los resultados para las siguientes configuraciones de censura:

	Diseño 1	Diseño 2	Diseño 3	Diseño 4	Diseño 5
$(X, 0) \approx$	75 %	55 %	40 %	55 %	40 %
$(X, 1) \approx$	20 %	40 %	55 %	20 %	40 %
$(X, 2) \approx$	5 %	5 %	5 %	25 %	20 %

Modelo de Cox para Datos Doblemente Censurados

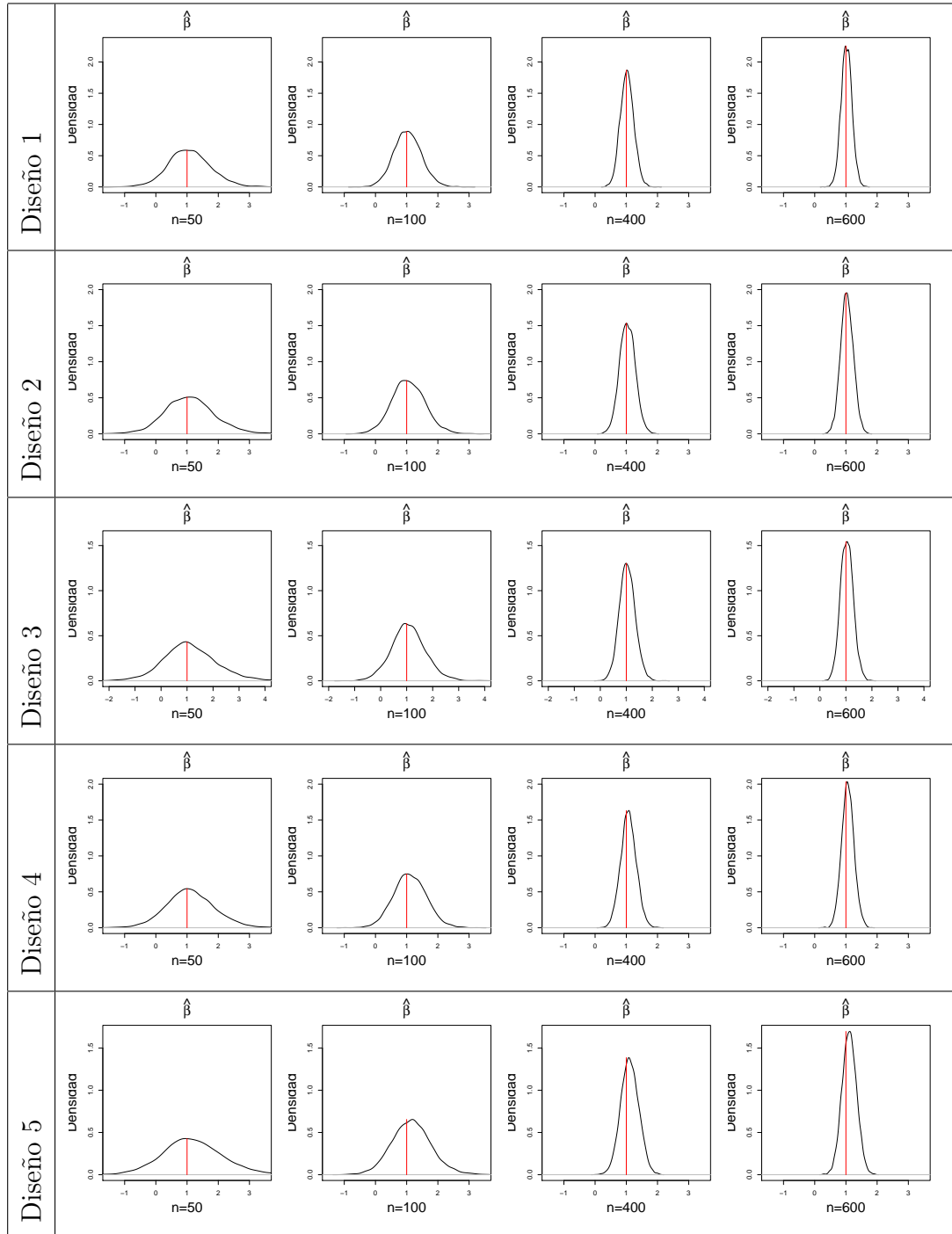


Figura 3.3: Densidad empírica de los parámetros estimados para 5000 réplicas con diferentes proporciones de censura y tamaños de muestra (una covariable).

Caso 2-dimensional

Para este caso, se considera $\beta_0 = (1, -1)'$ con $Z_1 \sim \mathcal{U}(0, 1)$ y $Z_2 \sim B(0, 4)$, se generan los tiempos de supervivencia T , se fija p y se generan los tiempos de censura C de tal modo que se obtienen las proporciones de datos completos, datos con censura por la izquierda y datos con censura por la derecha como sigue:

	Diseño 1	Diseño 2	Diseño 3
$(X, 0) \approx$	70 %	50 %	35 %
$(X, 1) \approx$	20 %	45 %	60 %
$(X, 2) \approx$	10 %	5 %	5 %

Los resultados para estos experimentos se muestran en la Tabla 3.6 y Figura 3.4.

Una vez generados los datos de supervivencia, se ajusta el modelo, es decir, se estiman los parámetros para cada base de datos y su error estándar (EE) en cada caso. En las Tablas 3.5 y 3.6 se muestran los resultados obtenidos para 5,000 réplicas, mostrando el promedio y la mediana de los parámetros estimados, así como el promedio del EE estimado y el EE empírico (EEe), además, del p -valor para el test de normalidad de Shapiro–Wilks, y la probabilidad empírica de cubrimiento (PC) para valores nominales de 90 %, 95 % y 99 %. Por otro lado, en las Figuras 3.3 y 3.4 se muestran las gráficas de la función de densidad estimada con una línea en el verdadero valor del parámetro β_0 para cada caso, que permite comprobar la normalidad de los estimadores para muestras finitas.

En las Tablas 3.5 y 3.6 se puede ver que el promedio y la mediana del valor obtenido para la estimación parámetro se acerca al verdadero valor cuando n crece. También se observa que el EE estimado es cercano al EE empírico y que éstos disminuyen cuando n crece. Por otro lado, el p -valor del test de normalidad indica que para valores grandes de n se aceptaría la hipótesis de normalidad con mayor o menor evidencia estadística, dependiendo no solo del valor de n , sino del diseño.

Modelo de Cox para Datos Doblemente Censurados

Dis	n	Param	Media	Mediana	EE	EEe	p-valor	PC para .90	PC para .95	PC para .99
1	50	β_1	1.09084	1.07127	0.65950	0.69524	0.00000	0.884	0.940	0.986
		β_2	-1.08253	-1.07104	0.41218	0.44421	0.00000	0.875	0.932	0.979
	100	β_1	1.05265	1.03226	0.44418	0.46068	0.00017	0.888	0.940	0.989
		β_2	-1.05519	-1.04980	0.27765	0.29018	0.00001	0.883	0.940	0.987
	400	β_1	1.02763	1.02593	0.21378	0.21628	0.00530	0.898	0.948	0.989
		β_2	-1.02781	-1.02519	0.13400	0.13646	0.00001	0.891	0.944	0.987
	600	β_1	1.02765	1.02433	0.17388	0.17601	0.32349	0.890	0.943	0.989
		β_2	-1.03020	-1.02995	0.10899	0.11062	0.18404	0.881	0.937	0.987
2	50	β_1	1.14087	1.10856	0.80945	0.89308	0.00000	0.869	0.925	0.981
		β_2	-1.13086	-1.09673	0.55690	0.64604	0.00000	0.865	0.924	0.971
	100	β_1	1.08266	1.07698	0.53275	0.56496	0.00000	0.882	0.938	0.985
		β_2	-1.07800	-1.05629	0.33404	0.35337	0.00000	0.880	0.933	0.982
	400	β_1	1.04135	1.04122	0.25382	0.26597	0.00486	0.885	0.941	0.986
		β_2	-1.04461	-1.03968	0.15895	0.15872	0.00002	0.886	0.938	0.986
	600	β_1	1.04053	1.03614	0.20605	0.20887	0.10063	0.894	0.943	0.986
		β_2	-1.04580	-1.04045	0.12927	0.13069	0.00000	0.875	0.933	0.984
3	50	β_1	1.21311	1.14997	1.06180	1.24826	0.00000	0.854	0.913	0.973
		β_2	-1.24325	-1.11557	1.15722	1.17604	0.00000	0.849	0.902	0.957
	100	β_1	1.09160	1.08532	0.66999	0.71852	0.00000	0.878	0.935	0.988
		β_2	-1.10686	-1.08411	0.42174	0.46088	0.00000	0.871	0.927	0.976
	400	β_1	1.05554	1.05011	0.31193	0.31911	0.00003	0.891	0.942	0.988
		β_2	-1.04905	-1.04664	0.19574	0.19806	0.00003	0.886	0.939	0.984
	600	β_1	1.04508	1.04410	0.25208	0.26127	0.25761	0.884	0.938	0.987
		β_2	-1.04775	-1.04473	0.15845	0.16709	0.11769	0.868	0.928	0.980

Tabla 3.6: Resultados para 5000 réplicas con diferentes proporciones de censura y diferentes tamaños de muestra (datos con dos covariables).

Modelo de Cox para Datos Doblemente Censurados

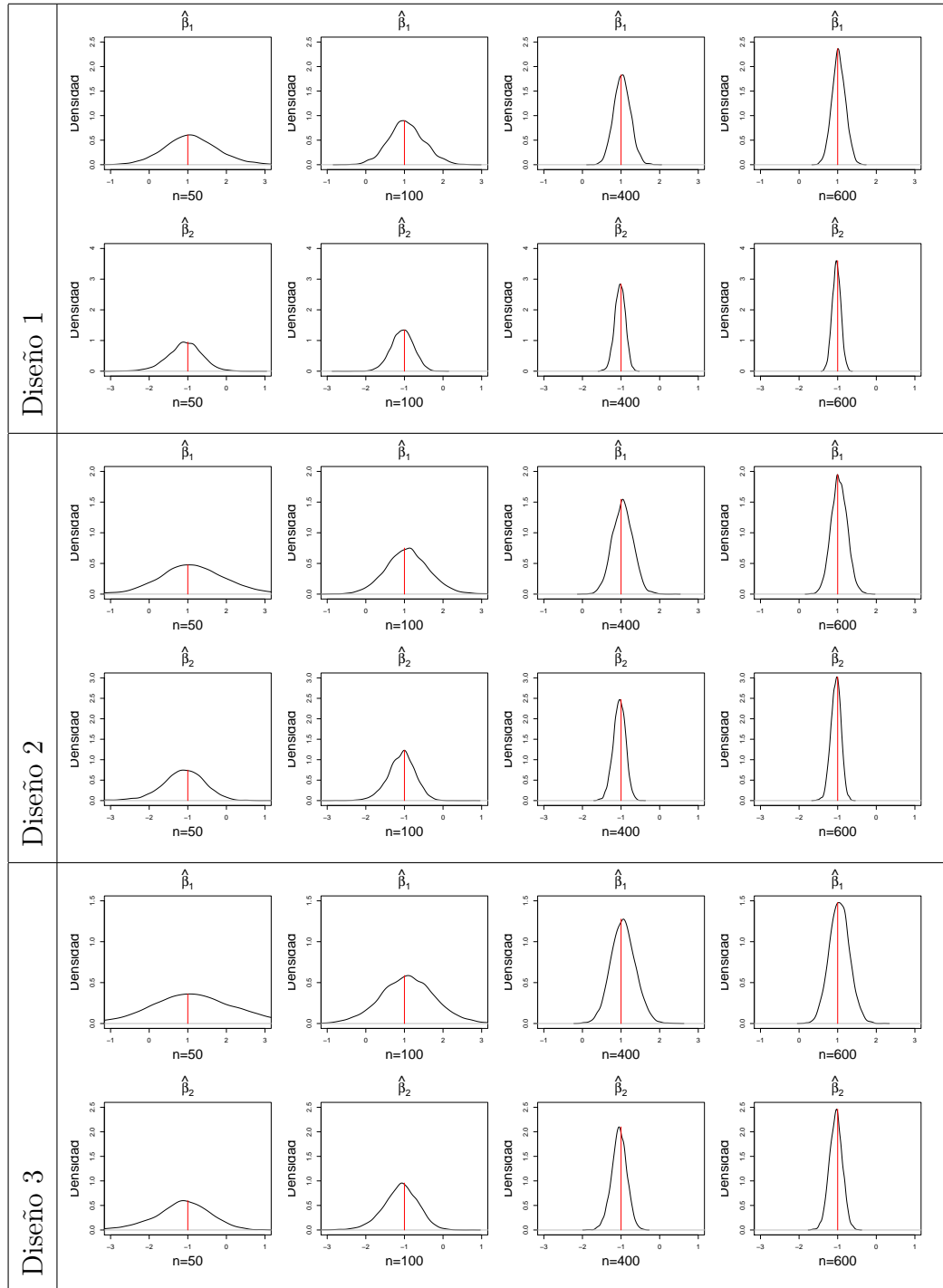


Figura 3.4: Densidad empírica de los parámetros estimados para 5000 réplicas con diferentes proporciones de censura y tamaños de muestra (dos covariables).

Por ejemplo, para datos con una sola covariable, Tabla 3.5, los estimadores de los parámetros en el Diseño 5 se distribuyen claramente como una normal para tamaños muestrales mayores ó iguales a 400. Sin embargo, los estimadores de los parámetros en el Diseño 3 parecen necesitar un tamaño muestral más grande para que el test de Shapiro–Wilks no rechace la hipótesis de normalidad. Es decir, que lo que parece tener más influencia es la proporción de datos censurados por la izquierda. Aunque, si comparamos los resultados para los Diseños 4 y 5 se aprecia que, a igualdad de datos censurados por la izquierda, en el diseño con un mayor número de datos censurados por la derecha éstos empeoran e incluso se ralentiza su aproximación a la distribución normal. En cuanto a datos con dos covariables, Tabla 3.6, los estimadores de los parámetros en el Diseño 3 se distribuyen claramente como una normal para tamaños muestrales mayores ó iguales a 600. Sin embargo, los estimadores de los parámetros en el Diseño 2 parecen necesitar un tamaño muestral más grande para que el test de Shapiro–Wilks no rechace la hipótesis de normalidad. Al igual que antes, cuantos menos datos censurados mejores resultados y a igualdad de datos censurados por la izquierda, el aumento del número de datos censurados por la derecha empeora los resultados. Por último, la probabilidad empírica de cubrimiento es bastante buena para todos los valores nominales, todos los tamaños muestrales e incluso no hay mucha diferencia para los diferentes diseños tanto para una como para dos covariables.

En las Figuras 3.3 y 3.4, se puede ver que las gráficas de la función de distribución empírica de los estimadores de los parámetros se aproximan a una distribución normal cuando n crece. Además, se puede ver que aunque en algunos casos el test de Shapiro–Wilks rechace la hipótesis de normalidad la forma de la densidad empírica es claramente la de una campana de Gauss. También se puede observar que el sesgo crece, con respecto al verdadero valor del parámetro, en configuraciones con menos datos completos.

Finalmente, en la Figura 3.5 se muestra la función de supervivencia estimada para el valor de la covariable $z = 0.5$ para 500 réplicas de los 5 diseños del caso 1-dimensional con tamaño de muestra $n = 100$ y 400 , comparada con la verdadera función de supervivencia. Obsérvese que cuando n crece la variabilidad de la estimación es menor. Nótese que los Diseños 3 y 5 tienen una banda ligeramente más ancha, sobretodo para $n = 100$, como era de esperar ya que estos diseños son los que tienen menor proporción de datos completos lo cual aumenta la variabilidad.

Modelo de Cox para Datos Doblemente Censurados

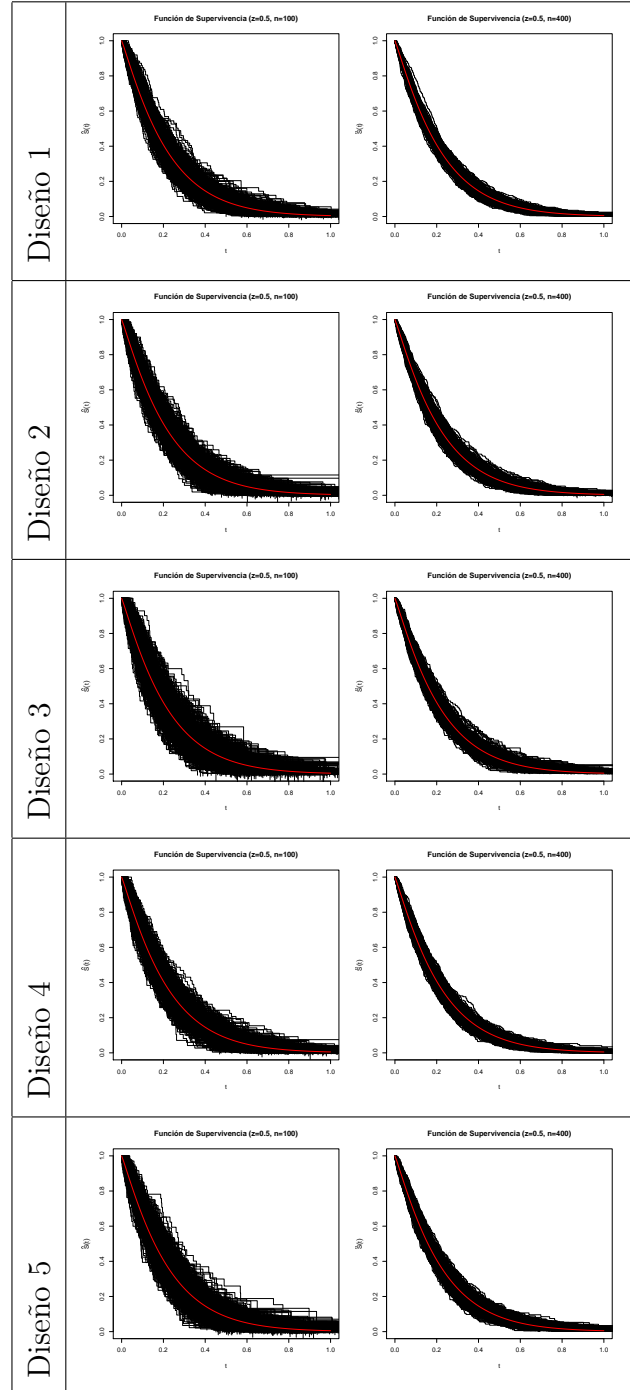


Figura 3.5: Gráfica de las funciones de supervivencia estimadas para 500 réplicas para los Diseños 1,2,3,4 y 5 con una covariable. En rojo se muestra la verdadera función de supervivencia.

Capítulo 4

Inferencia para los Modelos

4.1. Introducción

Una vez que se ha ajustado un modelo, es decir, se han estimado sus parámetros para un conjunto de datos de supervivencia dado, es necesario verificar el efecto de las covariables en éste. Algunos de los estadísticos clásicos para llevar a cabo este contraste en el modelo de Cox, son los estadísticos de *Wald* y cociente de verosimilitud (*LR*), estos estadísticos se pueden ver en Klein y Moeschberger (2003). En este capítulo, se proponen estadísticos alternativos basados en medidas de divergencia para hacer el contraste de hipótesis para los modelos estudiados en los capítulos anteriores. Además, se realiza un estudio de simulación para analizar el comportamiento de los estadísticos en cuanto a su potencia exacta y tamaño exacto, y compararlos con los estadísticos clásicos de Wald y cociente de verosimilitud.

Se consideran datos de supervivencia, estudiados en el Capítulo 3, doblemente censurados a los que se les ajusta un modelo semiparamétrico de riesgos proporcionales de Cox

$$\lambda(t|\mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z}) \quad (4.1.1)$$

y, datos de supervivencia, estudiados en el Capítulo 2, con censura por la derecha

a los que se les ajusta un modelo generalizado semiparamétrico de riesgos proporcionales

$$\lambda(t|\mathbf{z}) = \lambda_0(t)K(\boldsymbol{\beta}, a, \mathbf{z}), \quad (4.1.2)$$

donde, en ambos, $\lambda_0(\cdot)$ es la función de riesgo base, $\mathbf{z} \in \mathbb{R}^d$ es el vector de covariables y $\boldsymbol{\beta} \in \mathbb{R}^d$ es el vector de parámetros desconocidos. Además, para el modelo (4.1.2) se tiene el parámetro $a > 0$ con $K(\boldsymbol{\beta}, a, \mathbf{z})$ como en (2.2.2).

Nótese que las expresiones (4.1.1) y (4.1.2) se pueden escribir como:

$$\lambda(t|\mathbf{z}) = \lambda_0(t)\varphi(\mathbf{z}, \boldsymbol{\beta}) \quad (4.1.3)$$

con $\varphi(\mathbf{z}, \boldsymbol{\beta}) = \exp(\boldsymbol{\beta}'\mathbf{z})$ para la expresión (4.1.1) y $\varphi(\mathbf{z}, \boldsymbol{\beta}) = K(\boldsymbol{\beta}, a, \mathbf{z})$, donde a es el valor del parámetro estimado para el modelo (4.1.2). Entonces, la correspondiente función de densidad $f(\cdot)$ viene dada por

$$f(t|\mathbf{z}) = \lambda(t)S(t|\mathbf{z}) \quad (4.1.4)$$

donde $S(t|\mathbf{z}) = S_0(t)^{\varphi(\mathbf{z}, \boldsymbol{\beta})}$ con $S_0(\cdot)$ la función de supervivencia base.

Supóngase que se estiman los parámetros $\hat{\boldsymbol{\beta}}_n$ del modelo para una muestra de tamaño n de datos iid $(X_i, \delta_i, \mathbf{Z}_i)$ ó (X_i, A_i, \mathbf{Z}_i) , de acuerdo como se estudió en los Capítulos 2 y 3, respectivamente. Una vez hecho esto, es importante hacer un contraste de hipótesis sobre la significación del efecto del vector de covariables \mathbf{z} , es decir, la hipótesis nula es

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \quad (4.1.5)$$

frente a la hipótesis alternativa

$$H_1 : \boldsymbol{\beta} \neq \mathbf{0}.$$

Usando la expresión (4.1.4) se tiene que bajo H_1 , la función de densidad viene dada por

$$f(t|\mathbf{z}) = \lambda_0(t)\varphi(\mathbf{z}, \boldsymbol{\beta})S_0(t)^{\varphi(\mathbf{z}, \boldsymbol{\beta})} \quad (4.1.6)$$

mientras que la función de densidad bajo H_0 es

$$f_0(t) = C\lambda_0(t)S_0(t)^C \quad (4.1.7)$$

con $C = \varphi(\mathbf{z}, \mathbf{0})$. Obsérvese que para el modelo de Cox (4.1.1), $C = 1$ y para el modelo generalizado (4.1.2) se tiene $C = \frac{1}{2a}$.

A la hora de llevar a cabo el contraste de hipótesis dado en (4.1.5) para el modelo de Cox para datos con censura por la derecha, los estadísticos usuales son los estadísticos de Wald y del cociente de verosimilitudes que vienen dados por

$$Wald(\hat{\beta}_n) = n\hat{\beta}_n'\Sigma_0^{-1}\hat{\beta}_n$$

y

$$LR(\hat{\beta}_n) = 2[\tilde{\ell}_n(\hat{\beta}_n) - \tilde{\ell}_n(\mathbf{0})],$$

respectivamente, donde $\hat{\beta}_n$ es el estimador de máxima verosimilitud de los parámetros, Σ_0 es la matriz de covarianza del estimador y $\tilde{\ell}_n(\cdot)$ la función de verosimilitud del correspondiente modelo. Ambos estadísticos se pueden utilizar también para los modelos (4.1.1) y (4.1.2) con datos doblemente censurados y con censura por la derecha, respectivamente y se distribuyen asintóticamente como una distribución χ^2 con grados de libertad igual al número de parámetros desconocidos.

Sin embargo, en los últimos años, familias de estadísticos de contraste basados en medidas de divergencia se han considerado en diferentes contextos. Aunque el primer antecedente de la utilización de medidas de divergencia como base para definir estadísticos de contraste es el trabajo de Pearson (1900), ya que el estadístico de Pearson no es más que una medida de divergencia, es en los estudios de Cressie y Read (1984), Zografos (1993), Zografos y otros (1990), Pardo y otros (1993a,b), Morales y otros (1994a,b), Pardo (1994a,b), entre otros, en donde de una forma explícita, se hace referencia a la importancia de las medidas de divergencia en la definición de estadísticos de contraste para datos categorizables. No obstante, también hay algunos trabajos en este sentido en análisis de supervivencia. Chen

y otros (2004) proponen estadísticos de bondad de ajuste, basados en una familia de divergencias definida y estudiada por Read y Cressie (1988) y Cressie y Read (1984), para datos de supervivencia agrupados. Es más, no solo es posible construir estadísticos basados en medidas de divergencia, sino también estimadores. Yang (1991) estudió la robustez del estimador de Hellinger para datos de supervivencia continuos. Recientemente, Devarajan y Ebrahimi (2009) presentaron estadísticos para contrastar el efecto del vector de covariables en el modelo de Cox para datos con censura por la derecha, basados en diferentes medidas de divergencia.

En la Sección 4.2, se definen varios estadísticos basados en medidas de divergencia para llevar a cabo el contraste de hipótesis dado en (4.1.5), tanto para el modelo de Cox (4.1.1) con datos doblemente censurados, como para el modelo semiparamétrico generalizado de riesgos proporcionales (4.1.2) con datos con censura por la derecha. Además, se calcula la distribución asintótica de estos estadísticos bajo la hipótesis (4.1.5) en la Sección 4.3, lo que permitirá construir la región crítica correspondiente para llevar a cabo el respectivo contraste de hipótesis. En la Sección 4.4, se calcula la potencia asintótica de estos estadísticos. La distribución asintótica de los estadísticos introducidos es independiente de la divergencia utilizada en la construcción específica del estadístico, es decir, son asintóticamente equivalentes. Sin embargo, eso no implica, obviamente, que su comportamiento sea el mismo para tamaños muestrales pequeños o moderados. De esta forma, para comparar los estadísticos, se lleva a cabo un estudio de simulación en la Sección 4.5.

4.2. Estadísticos Basados en Divergencias

Una *medida de divergencia* es una medida de discrepancia entre dos funciones de distribución de tal modo que

- $D(F : F_0) \geq 0$ para $F, F_0 \in \mathcal{F}$,

- $D(F : F_0) = 0$ si y sólo si $F = F_0$,

donde \mathcal{F} es el espacio de distribuciones de probabilidad con soporte común.

La divergencia de Kullback–Leibler (KL) entre las funciones de distribución $F(\cdot)$ y $F_0(\cdot)$ con dirección $F(\cdot)$ está dada por

$$KL(F : F_0) = \int_0^\infty f(x) \log \frac{f(x)}{f_0(x)} dx \quad (4.2.8)$$

donde $f(\cdot)$ y $f_0(\cdot)$ son las funciones de densidad de $F(\cdot)$ y $F_0(\cdot)$, respectivamente. Esta divergencia se denomina “minimum discrimination information” entre $F(\cdot)$ y $F_0(\cdot)$ y fue introducida y estudiada por Kullback y Leibler (1951) y Kullback (1959). De forma similar, la divergencia KL entre $F(\cdot)$ y $F_0(\cdot)$ con dirección $F_0(\cdot)$ es

$$KL(F_0 : F) = \int_0^\infty f_0(x) \log \frac{f_0(x)}{f(x)} dx. \quad (4.2.9)$$

La J-divergencia combina las divergencias (4.2.8) y (4.2.9) y se define como

$$J(F : F_0) = KL(F : F_0) + KL(F_0 : F) \quad (4.2.10)$$

y fue introducida por Jeffreys (1946).

Una medida que generaliza la divergencia KL es la divergencia de Rényi (R). La divergencia de Rényi de orden γ entre las distribuciones $F(\cdot)$ y $F_0(\cdot)$ con densidades $f(\cdot)$ y $f_0(\cdot)$ vienen dada por

$$R_\gamma(F_0 : F) = \frac{1}{\gamma - 1} \log \left[\int_0^\infty \left(\frac{f_0(x)}{f(x)} \right)^{\gamma-1} f_0(x) dx \right] \quad (4.2.11)$$

con $\gamma \neq 1$. El caso $\gamma = 1$ se define como $R_1(F_0 : F) = \lim_{\gamma \rightarrow 1} R_\gamma(F_0 : F) = KL(F : F_0)$. Si $\gamma = \frac{1}{2}$ la divergencia es simétrica y fue introducida por Bhattacharyya (1943).

Ahora, sustituyendo las expresiones de las funciones de densidad (4.1.6) y (4.1.7) en la divergencia KL (4.2.8) se tiene

$$\begin{aligned}
 KL(F : F_0|\boldsymbol{\beta}) &= \int_0^\infty f(t|\mathbf{z}) \log \left(\frac{f(t|\mathbf{z})}{f_0(t)} \right) dt \\
 &= \int_0^\infty f(t|\mathbf{z}) \log \left(\frac{\lambda_0(t)\varphi(\mathbf{z}, \boldsymbol{\beta})S_0(t)^{\varphi(\mathbf{z}, \boldsymbol{\beta})}}{C\lambda_0(t)S_0(t)^C} \right) dt \\
 &= \int_0^\infty f(t|\mathbf{z}) \log \left(\frac{\varphi(\mathbf{z}, \boldsymbol{\beta})S_0(t)^{\varphi(\mathbf{z}, \boldsymbol{\beta})-C}}{C} \right) dt \\
 &= \int_0^\infty f(t|\mathbf{z}) \left[\log \left(\frac{\varphi(\mathbf{z}, \boldsymbol{\beta})}{C} \right) + \log (S_0(t)^{\varphi(\mathbf{z}, \boldsymbol{\beta})-C}) \right] dt \\
 &= \log \left(\frac{\varphi(\mathbf{z}, \boldsymbol{\beta})}{C} \right) \int_0^\infty f(t|\mathbf{z}) dt + (\varphi(\mathbf{z}, \boldsymbol{\beta}) - C) \int_0^\infty f(t|\mathbf{z}) \log S_0(t) dt \\
 &= \log \left(\frac{\varphi(\mathbf{z}, \boldsymbol{\beta})}{C} \right) - \frac{(\varphi(\mathbf{z}, \boldsymbol{\beta}) - C)}{\varphi(\mathbf{z}, \boldsymbol{\beta})} \int_0^\infty -f(t|\mathbf{z})\varphi(\mathbf{z}, \boldsymbol{\beta}) \log S_0(t) dt \\
 &= \log \left(\frac{\varphi(\mathbf{z}, \boldsymbol{\beta})}{C} \right) - \frac{\varphi(\mathbf{z}, \boldsymbol{\beta}) - C}{\varphi(\mathbf{z}, \boldsymbol{\beta})} \int_0^\infty f(t|\mathbf{z}) (-\log S_0(t)^{\varphi(\mathbf{z}, \boldsymbol{\beta})}) dt \\
 &= \log \left(\frac{\varphi(\mathbf{z}, \boldsymbol{\beta})}{C} \right) - \left(1 - \frac{C}{\varphi(\mathbf{z}, \boldsymbol{\beta})} \right) \int_0^\infty f(t|\mathbf{z}) \Lambda(t|\mathbf{z}) dt \\
 &= \log \left(\frac{\varphi(\mathbf{z}, \boldsymbol{\beta})}{C} \right) + \left(\frac{C}{\varphi(\mathbf{z}, \boldsymbol{\beta})} - 1 \right) \int_0^\infty f(t|\mathbf{z}) \Lambda(t|\mathbf{z}) dt \quad (4.2.12)
 \end{aligned}$$

donde $\Lambda(\cdot)$ es la función de riesgo acumulativa.

Por otro lado, integrando por partes se tiene,

$$\begin{aligned}
 \int_0^\infty f(t|\mathbf{z}) \Lambda(t|\mathbf{z}) dt &= \Lambda(t|\mathbf{z}) F(t|\mathbf{z})|_0^\infty - \int_0^\infty F(t|\mathbf{z}) \lambda(t|\mathbf{z}) dt \\
 &= \Lambda(t|\mathbf{z}) F(t|\mathbf{z})|_0^\infty - \int_0^\infty (1 - S(t|\mathbf{z})) \lambda(t|\mathbf{z}) dt \\
 &= \Lambda(t|\mathbf{z}) F(t|\mathbf{z})|_0^\infty - \int_0^\infty \lambda(t|\mathbf{z}) dt + \int_0^\infty S(t|\mathbf{z}) \lambda(t|\mathbf{z}) dt \\
 &= \Lambda(t|\mathbf{z}) F(t|\mathbf{z})|_0^\infty - \Lambda(t|\mathbf{z})|_0^\infty + \int_0^\infty f(t|\mathbf{z}) dt \\
 &= \Lambda(t|\mathbf{z}) (F(t|\mathbf{z}) - 1)|_0^\infty + 1 \\
 &= -\Lambda(t|\mathbf{z}) S(t|\mathbf{z})|_0^\infty + 1 \\
 &= \log (S(t|\mathbf{z})) S(t|\mathbf{z})|_0^\infty + 1. \quad (4.2.13)
 \end{aligned}$$

Se sabe que la función de supervivencia $S(\cdot)$ cumple

$$\lim_{t \rightarrow \infty} S(t|\mathbf{z}) = 0$$

y

$$\lim_{t \rightarrow 0} S(t|\mathbf{z}) = 1$$

entonces, aplicando la regla de L'Hopital

$$\begin{aligned} \lim_{t \rightarrow \infty} \log(S(t|\mathbf{z})) S(t|\mathbf{z}) &= \lim_{t \rightarrow \infty} \frac{\frac{\log(S(t|\mathbf{z}))}{S(t|\mathbf{z})}}{\frac{1}{S(t|\mathbf{z})}} \\ &= \lim_{t \rightarrow \infty} \frac{\frac{S'(t|\mathbf{z})}{S(t|\mathbf{z})}}{\frac{-S'(t|\mathbf{z})}{(S(t|\mathbf{z}))^2}} \\ &= \lim_{t \rightarrow \infty} (-S(t|\mathbf{z})) \\ &= 0. \end{aligned}$$

Además,

$$\lim_{t \rightarrow 0} \log(S(t|\mathbf{z})) S(t|\mathbf{z}) = 0,$$

por tanto, de (4.2.13) se tiene

$$\int_0^\infty f(t|\mathbf{z}) \Lambda(t|\mathbf{z}) = 1$$

y por (4.2.12) se obtiene

$$KL(F : F_0|\boldsymbol{\beta}) = \log\left(\frac{\varphi(\mathbf{z}, \boldsymbol{\beta})}{C}\right) + \frac{C}{\varphi(\mathbf{z}, \boldsymbol{\beta})} - 1. \quad (4.2.14)$$

Análogamente, también se puede obtener que

$$KL(F_0 : F|\boldsymbol{\beta}) = \log\left(\frac{C}{\varphi(\mathbf{z}, \boldsymbol{\beta})}\right) + \frac{\varphi(\mathbf{z}, \boldsymbol{\beta})}{C} - 1. \quad (4.2.15)$$

Usando (4.2.10) y las expresiones anteriores, se obtiene la divergencia de Jeffrey entre las funciones de densidad (4.1.6) y (4.1.7) que viene dada por

$$\begin{aligned} J(F : F_0|\boldsymbol{\beta}) &= \log \left(\frac{\varphi(\mathbf{z}, \boldsymbol{\beta})}{C} \right) + \frac{C}{\varphi(\mathbf{z}, \boldsymbol{\beta})} - 1 + \log \left(\frac{C}{\varphi(\mathbf{z}, \boldsymbol{\beta})} \right) + \frac{\varphi(\mathbf{z}, \boldsymbol{\beta})}{C} - 1 \\ &= \frac{C}{\varphi(\mathbf{z}, \boldsymbol{\beta})} + \frac{\varphi(\mathbf{z}, \boldsymbol{\beta})}{C} - 2. \end{aligned} \quad (4.2.16)$$

Finalmente, para obtener la divergencia de Rényi entre las funciones de densidad (4.1.6) y (4.1.7) se tiene

$$\begin{aligned} \int_0^\infty \left(\frac{f_0(t)}{f(t|\mathbf{z})} \right)^{\gamma-1} f_0(t) dt &= \int_0^\infty f_0(t)^\gamma f(t|\mathbf{z})^{1-\gamma} dt \\ &= \int_0^\infty (\lambda_0(t) C S_0(t)^C)^\gamma (\lambda_0(t) \varphi(\mathbf{z}, \boldsymbol{\beta}) S_0(t)^{\varphi(\mathbf{z}, \boldsymbol{\beta})})^{1-\gamma} dt \\ &= \int_0^\infty C^\gamma (\varphi(\mathbf{z}, \boldsymbol{\beta}))^{1-\gamma} \lambda_0(t) S_0(t)^{\varphi(\mathbf{z}, \boldsymbol{\beta})(1-\gamma)+C\gamma} dt \\ &= c_1 \int_0^\infty \lambda_0(t) S_0(t)^{c_2} dt, \end{aligned}$$

donde

$$c_1 = C^\gamma (\varphi(\mathbf{z}, \boldsymbol{\beta}))^{1-\gamma} \text{ y } c_2 = (1 - \gamma)\varphi(\mathbf{z}, \boldsymbol{\beta}) + \gamma C, \quad (4.2.17)$$

por tanto,

$$\begin{aligned}
 c_1 \int_0^\infty \lambda_0(t) S_0(t)^{c_2} dt &= c_1 \int_0^\infty \lambda_0(t) \exp(\log S_0(t)^{c_2}) dt \\
 &= c_1 \int_0^\infty \lambda_0(t) \exp\left(\frac{c_2}{C} \log S_0(t)^C\right) dt \\
 &= c_1 \int_0^\infty \lambda_0(t) \exp\left(-\frac{c_2}{C} \Lambda_0(t)\right) dt \\
 &= -\frac{c_1}{c_2} \int_0^\infty -\frac{c_2}{C} \lambda_0(t) C \exp\left(-\frac{c_2}{C} \Lambda_0(t)\right) dt \\
 &= -\frac{c_1}{c_2} \exp\left(-\frac{c_2}{C} \Lambda_0(t)\right) \Big|_0^\infty \\
 &= -\frac{c_1}{c_2} \exp(\log S_0(t)^{c_2}) \Big|_0^\infty \\
 &= -\frac{c_1}{c_2} S_0(t)^{c_2} \Big|_0^\infty \\
 &= -\frac{c_1}{c_2} (0 - 1) \\
 &= \frac{c_1}{c_2}.
 \end{aligned}$$

Entonces, usando la expresión (4.2.11) y (4.2.17) se tiene

$$\begin{aligned}
 R_\gamma(F_0 : F|\boldsymbol{\beta}) &= \frac{1}{\gamma - 1} \log \frac{c_1}{c_2} \\
 &= \frac{1}{\gamma - 1} (\log c_1 - \log c_2) \\
 &= \frac{1}{\gamma - 1} [\log (C^\gamma \varphi(\mathbf{z}, \boldsymbol{\beta})^{1-\gamma}) - \log ((1 - \gamma)\varphi(\mathbf{z}, \boldsymbol{\beta}) + \gamma C)] \\
 &= \frac{1}{1 - \gamma} [\log ((1 - \gamma)\varphi(\mathbf{z}, \boldsymbol{\beta}) + \gamma C) - \gamma \log C] - \log \varphi(\mathbf{z}, \boldsymbol{\beta}).
 \end{aligned} \tag{4.2.18}$$

Ahora, supóngase que $\hat{\boldsymbol{\beta}}_n$ es el parámetro estimado del correspondiente modelo generalizado de riesgos proporcionales o del modelo de Cox, en base a una muestra iid de tamaño n con datos de la forma $(X_i, \delta_i, \mathbf{Z}_i)$, es decir, datos con censura por la derecha ó con datos de la forma (X_i, A_i, \mathbf{Z}_i) , es decir, datos doblemente censurados, respectivamente. Para hacer el contraste de hipótesis usando las divergencias

(4.2.14), (4.2.15), (4.2.16) y (4.2.18) se proponen los siguientes estadísticos

$$\begin{aligned} KL_1(\hat{\beta}_n) &= 2nKL((F : F_0|\hat{\beta}_n) \\ &= 2n \sum_{i=1}^n \left[\log \left(\frac{\varphi(\mathbf{z}_i, \hat{\beta}_n)}{C} \right) - 1 + \frac{C}{\varphi(\mathbf{z}_i, \hat{\beta}_n)} \right], \end{aligned} \quad (4.2.19)$$

$$\begin{aligned} KL_2(\hat{\beta}_n) &= 2nKL(F_0 : F|\hat{\beta}_n) \\ &= 2n \sum_{i=1}^n \left[\log \left(\frac{C}{\varphi(\mathbf{z}_i, \hat{\beta}_n)} \right) - 1 + \frac{\varphi(\mathbf{z}_i, \hat{\beta}_n)}{C} \right], \end{aligned} \quad (4.2.20)$$

$$\begin{aligned} \tilde{J}(\hat{\beta}_n) &= 2nJ(F : F_0|\hat{\beta}_n) \\ &= 2n \sum_{i=1}^n \left[\frac{C}{\varphi(\mathbf{z}_i, \hat{\beta}_n)} + \frac{\varphi(\mathbf{z}_i, \hat{\beta}_n)}{C} - 2 \right] \end{aligned} \quad (4.2.21)$$

y

$$\begin{aligned} \tilde{R}_\gamma(\hat{\beta}_n) &= 2nR_\gamma(F_0 : F|\hat{\beta}_n) \\ &= 2n \frac{1}{1-\gamma} \sum_{i=1}^n \log \left[(1-\gamma)\varphi(\mathbf{z}_i, \hat{\beta}_n) + \gamma C \right] - \sum_{i=1}^n \log \varphi(\mathbf{z}_i, \hat{\beta}_n) - \frac{n\gamma \log C}{1-\gamma}, \end{aligned} \quad (4.2.22)$$

respectivamente.

4.3. Distribución Asintótica de los Estadísticos

En el siguiente teorema se establece la distribución asintótica de los estadísticos (4.2.19)–(4.2.22). Sin embargo, antes de demostrar dicho teorema es necesario tener en cuenta el siguiente corolario.

Corolario 4.3.1. *Nótese que para el modelo generalizado con datos con censura por la derecha (4.1.2), bajo H_0 y usando el resultado del Teorema 2.6.8 se tiene*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \sim \mathcal{N}(\mathbf{0}, \Sigma_0^*),$$

con $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\beta}}'_n, \hat{a})'$, $\boldsymbol{\theta}_0 = (\mathbf{0}, a)$ y $\boldsymbol{\Sigma}_0^*$ la matriz de covarianza del vector $\hat{\boldsymbol{\theta}}_n$. De aquí se tiene que

$$\sqrt{n}\hat{\boldsymbol{\beta}}_n \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_0),$$

con $\boldsymbol{\Sigma}_0$ la matriz de covarianza del vector $\hat{\boldsymbol{\beta}}_n$.

Teorema 4.3.2. Sea $\hat{\boldsymbol{\beta}}_n$ el estimador de máxima verosimilitud del modelo (4.1.3) con el respectivo mecanismo de censura, además, se asumen las condiciones del Corolario 4.3.1 y del Teorema 3.4.7, respectivamente. Bajo H_0 definido en (4.1.5), los estadísticos (4.2.19)-(4.2.22) se distribuyen asintóticamente como

$$\sum_{i=1}^r \lambda_i \chi_{1i}^2,$$

donde $r = \text{rango}(\boldsymbol{\Sigma}_0 \nabla^2 T_n(\mathbf{0}) \boldsymbol{\Sigma}_0)$, χ_{1i}^2 para $i = 1, \dots, r$ son variables aleatorias independientes ji-cuadrado con un grado de libertad y $\lambda_1, \dots, \lambda_r$ son los autovalores diferentes de cero de la matriz $\frac{1}{2} \nabla^2 T_n(\mathbf{0}) \boldsymbol{\Sigma}_0$, con $T_n(\cdot)$ el respectivo estadístico y $\boldsymbol{\Sigma}_0$ la matriz de covarianza del estimador $\hat{\boldsymbol{\beta}}_n$ bajo H_0 .

Demostración. Para el estadístico $KL_1(\hat{\boldsymbol{\beta}}_n)$ se considera su desarrollo en serie de Taylor de segundo orden alrededor de $\mathbf{0}$, entonces, $KL_1(\hat{\boldsymbol{\beta}}_n)$ se puede aproximar por

$$\begin{aligned} KL_1(\hat{\boldsymbol{\beta}}_n) &\approx KL_1(\mathbf{0}) + (\hat{\boldsymbol{\beta}}_n - \mathbf{0})' \nabla KL_1(\mathbf{0}) \\ &\quad + \frac{1}{2!} (\hat{\boldsymbol{\beta}}_n - \mathbf{0})' \nabla^2 KL_1(\mathbf{0}) (\hat{\boldsymbol{\beta}}_n - \mathbf{0}) \end{aligned} \quad (4.3.23)$$

donde $\nabla KL_1(\mathbf{0})$ y $\nabla^2 KL_1(\mathbf{0})$ son el vector gradiente y la matriz hessiana de $KL_1(\cdot)$ evaluado en $\mathbf{0}$, respectivamente.

Como $\varphi(\mathbf{z}_i, \mathbf{0}) = C$, se tiene

$$KL_1(\mathbf{0}) = 2n \sum_{i=1}^n \left[\ln \left(\frac{C}{C} \right) - 1 + \frac{C}{C} \right] = 0.$$

Además,

$$\begin{aligned}\nabla KL_1(\mathbf{0}) &= 2n \sum_{i=1}^n \left[\frac{C}{\varphi(\mathbf{z}_i, \mathbf{0})} \frac{\varphi'(\mathbf{z}_i, \mathbf{0})}{C} - C \frac{\varphi'(\mathbf{z}_i, \mathbf{0})}{\varphi(\mathbf{z}_i, \mathbf{0})^2} \right] \\ &= 2n \sum_{i=1}^n \left[\frac{\varphi'(\mathbf{z}_i, \mathbf{0})}{C} - \frac{\varphi'(\mathbf{z}_i, \mathbf{0})}{C} \right] \\ &= 0.\end{aligned}$$

Entonces, usando la serie (4.3.23) se tiene

$$\begin{aligned}KL_1(\hat{\beta}_n) &\approx \frac{1}{2} \hat{\beta}_n' \nabla^2 KL_1(\mathbf{0}) \hat{\beta}_n, \\ &= \frac{1}{2} \hat{\beta}_n' 2n \nabla^2 KL(F : F_0 | \mathbf{0}) \hat{\beta}_n, \\ &= n \hat{\beta}_n' \nabla^2 KL(F : F_0 | \mathbf{0}) \hat{\beta}_n\end{aligned}\tag{4.3.24}$$

Por otro lado, teniendo en cuenta el Corolario 4.3.1 y el Teorema 3.4.7, se sabe que bajo H_0 se tiene que $\sqrt{n} \hat{\beta}_n$ se distribuye asintóticamente como $\mathcal{N}(\mathbf{0}, \Sigma_0)$, donde Σ_0 es la matriz de covarianza del estimador $\hat{\beta}_n$ bajo H_0 . Por el Corolario 2.1 de Dik y Gunst (1985) se tiene que

$$KL_1(\hat{\beta}_n) \xrightarrow[n \rightarrow \infty]{d} \sum_{i=1}^r \lambda_i \chi_{1i}^2$$

donde $r = \text{rango}(\Sigma_0 \nabla^2 KL_1(\mathbf{0}) \Sigma_0)$, $\lambda_1, \dots, \lambda_r$ son los autovalores diferentes de cero de la matriz $\frac{1}{2} \nabla^2 KL(\mathbf{0}) \Sigma_0$ y χ_{1i}^2 son variables aleatorias ji-cuadrado con un grado de libertad independientes.

Análogamente, se obtiene la distribución asintótica de los estadísticos (4.2.20)–(4.2.22). □

Corolario 4.3.3. *Para cualquiera de los estadísticos definidos en (4.2.19)–(4.2.22), denotado por T_n , se rechaza la hipótesis nula H_0 , con nivel de significación α , si $T_n > t_\alpha$, con t_α tal que $P(\sum_{i=1}^r \lambda_i \chi_{1i}^2 > t_\alpha) = \alpha$.*

De acuerdo al corolario anterior, para hacer el contraste de hipótesis es necesario calcular percentiles de una combinación lineal $T = \sum_{i=1}^r \lambda_i \chi_{1i}^2$. En la práctica, para r pequeña es posible considerar tablas (como en Solomon (1960), Johnson y Kotz (1968), Eckler (1969) y Gupta (1963)) de la distribución acumulativa de la expresión. Otra alternativa es que la distribución de T se puede aproximar por una distribución χ^2 con r grados de libertad. Por ejemplo, como Rao y Scott (1981) se puede tratar $\tilde{T} = \bar{\lambda}^{-1}T$ como una variable aleatoria χ^2 con r grados de libertad donde $\bar{\lambda}$ es la media de los valores λ_i . En este caso, la media asintótica de \tilde{T} coincide con la media de una distribución χ_r^2 , pero la varianza asintótica de \tilde{T} es igual a $2r(1 + m^2)$, donde m es el coeficiente de variación de los autovalores, esto es $m^2 = r^{-1}\bar{\lambda}^{-2} \sum_i (\lambda_i - \bar{\lambda})^2$. Otra aproximación se puede hacer como Satterthwaite (1946) que usa $\tilde{T} = [(1 + m^2)\bar{\lambda}]^{-1}T$ que se distribuye como una $\mathcal{G}(1/2, 1/2\nu)$ con $\nu = r(1 + m^2)^{-1}$, donde $\mathcal{G}(a, b)$ es una distribución Gamma con parámetros a y b .

No obstante, en el caso 1-dimensional la distribución asintótica de los estadísticos (4.2.19)–(4.2.22) es una distribución χ_1^2 como se establece en el siguiente corolario.

Corolario 4.3.4. *En el caso donde el modelo (4.1.4) tiene una sola covariable, bajo H_0 , los estadísticos*

$$Kullback_1^{(1)}(\hat{\beta}_n) := \frac{2C^2 KL(F : F_0 | \hat{\beta}_n)}{\sigma_0^2 \sum_{i=1}^n [\varphi'(z_i, 0)]^2},$$

$$Kullback_2^{(1)}(\hat{\beta}_n) := \frac{2C^2 KL(F_0 : F | \hat{\beta}_n)}{\sigma_0^2 \sum_{i=1}^n [\varphi'(z_i, 0)]^2},$$

$$Jeffreys^{(1)}(\hat{\beta}_n) := \frac{C^2 J(F : F_0 | \hat{\beta}_n)}{\sigma_0^2 \sum_{i=1}^n [\varphi'(z_i, 0)]^2}$$

y

$$Renyi_\gamma^{(1)}(\hat{\beta}_n) := \frac{2C^2 R_\gamma(F_0 : F | \hat{\beta}_n)}{\sigma_0^2 \gamma \sum_{i=1}^n [\varphi'(z_i, 0)]^2}$$

se distribuyen asintóticamente como una distribución χ_1^2 .

Demostración. En el caso 1-dimensional la función KL_1 de la expresión (4.3.24) viene dada por

$$KL_1(\hat{\beta}_n) \approx \frac{1}{2} \hat{\beta}_n \frac{\partial^2 KL_1(\beta)}{\partial \beta^2} \Big|_{\beta=0} \hat{\beta}_n. \quad (4.3.25)$$

Al ser

$$\begin{aligned} \frac{\partial^2 KL_1(\beta)}{\partial \beta^2} \Big|_{\beta=0} &= 2n \frac{\partial^2}{\partial \beta^2} \sum_{i=1}^n \left(\log \frac{\varphi(z_i, \beta)}{C} - 1 + \frac{C}{\varphi(z_i, \beta)} \right) \Big|_{\beta=0} \\ &= 2n \frac{\partial}{\partial \beta} \sum_{i=1}^n \left(\frac{\varphi'(z_i, \beta)}{\varphi(z_i, \beta)} - C \frac{\varphi'(z_i, \beta)}{\varphi(z_i, \beta)^2} \right) \Big|_{\beta=0} \\ &= 2n \sum_{i=1}^n \left(\frac{\varphi(z_i, \beta) \varphi''(z_i, \beta) - \varphi'(z_i, \beta) \varphi'(z_i, \beta)}{\varphi(z_i, \beta)^2} \right. \\ &\quad \left. - C \frac{\varphi(z_i, \beta)^2 \varphi''(z_i, \beta) - \varphi'(z_i, \beta) 2\varphi(z_i, \beta) \varphi'(z_i, \beta)}{\varphi(z_i, \beta)^4} \right) \Big|_{\beta=0} \\ &= 2n \sum_{i=1}^n \left(\frac{C}{C^2} \varphi''(z_i, 0) - \frac{\varphi'(z_i, 0) \varphi'(z_i, 0)}{C^2} \right. \\ &\quad \left. - \frac{C^3}{C^4} \varphi''(z_i, 0) + \frac{2C^2 \varphi'(z_i, 0) \varphi'(z_i, 0)}{C^4} \right) \\ &= 2n \sum_{i=1}^n \frac{\varphi'(z_i, 0) \varphi'(z_i, 0)}{C^2} \\ &= \frac{2n}{C^2} \sum_{i=1}^n [\varphi'(z_i, 0)]^2, \end{aligned}$$

entonces, usando la expresión (4.3.25) se obtiene

$$\begin{aligned} \frac{C^2}{2n \sum_{i=1}^n [\varphi'(z_i, 0)]^2} KL_1(\hat{\beta}_n) &\approx \hat{\beta}_n^2 \\ \frac{C^2}{n \sum_{i=1}^n [\varphi'(z_i, 0)]^2} 2n KL(F : F_0 | \hat{\beta}_n) &\approx \hat{\beta}_n^2. \end{aligned}$$

Por tanto, si se define el estadístico

$$Kullback_1^{(1)}(\hat{\beta}_n) := \frac{2C^2 KL(F : F_0 | \hat{\beta}_n)}{\sigma_0^2 \sum_{i=1}^n [\varphi'(z_i, 0)]^2}$$

como por el Corolario 4.3.1 y el Teorema 3.4.7 se tiene, asintóticamente, que bajo H_0 $\sqrt{n}\hat{\beta}_n \sim \mathcal{N}(0, \sigma_0^2)$ entonces $\hat{\beta}_n^2 \sim \chi_1^2$, por tanto,

$$Kullback_1^{(1)}(\hat{\beta}_n) \sim \chi_1^2.$$

De forma similar, se pueden definir los estadísticos $Kullback_2^{(1)}$, $Jeffreys^{(1)}$ y $Renyi_\gamma^{(1)}$ como

$$Kullback_2^{(1)}(\hat{\beta}_n) := \frac{2C^2 KL(F_0 : F|\hat{\beta}_n)}{\sigma_0^2 \sum_{i=1}^n [\varphi'(z_i, 0)]^2}$$

$$Jeffreys^{(1)}(\hat{\beta}_n) := \frac{C^2 J(F : F_0|\hat{\beta}_n)}{\sigma_0^2 \sum_{i=1}^n [\varphi'(z_i, 0)]^2}$$

y

$$Renyi_\gamma^{(1)}(\hat{\beta}_n) := \frac{2C^2 R_\gamma(F_0 : F|\hat{\beta}_n)}{\sigma_0^2 \gamma \sum_{i=1}^n [\varphi'(z_i, 0)]^2}.$$

Puesto que

$$\left. \frac{\partial^2}{\partial \beta^2} KL(F_0 : F|\beta) \right|_{\beta=0} = \frac{1}{C^2} \sum_{i=1}^n [\varphi'(z_i, 0)]^2$$

$$\left. \frac{\partial^2}{\partial \beta^2} J(F : F_0|\beta) \right|_{\beta=0} = \frac{2}{C^2} \sum_{i=1}^n [\varphi'(z_i, 0)]^2$$

y

$$\left. \frac{\partial^2}{\partial \beta^2} R_\gamma(F_0 : F|\beta) \right|_{\beta=0} = \frac{\gamma}{C^2} \sum_{i=1}^n [\varphi'(z_i, 0)]^2,$$

y por el argumento anterior, se tiene que asintóticamente se distribuyen como una χ^2 con un grado de libertad. □

El Teorema 4.3.2 nos permite hacer contrastes de hipótesis globales, sin embargo, algunas veces es necesario hacer contrastes de hipótesis para un subconjunto de los parámetros β . Si se particiona el vector de parámetros $\beta = (\beta^{(1)}, \beta^{(2)})$, donde $\beta^{(1)}$ es el subvector $q \times 1$ de los primeros q componentes de β y $\beta^{(2)}$ es el subvector restante de tamaño $(d - q) \times 1$, la hipótesis nula de interés es

$$H_0 : \beta^{(1)} = \beta_0^{(1)}. \tag{4.3.26}$$

La función de densidad del modelo bajo esta hipótesis se puede expresar como

$$f_0(t) = C_s \lambda_0(t) S_0(t)^{C_s},$$

con $C_s = \varphi(z, (\beta_0^{(1)}, \beta^{(2)}))$. Entonces las expresiones de los estadísticos (4.2.19)–(4.2.22) siguen siendo válidas sustituyendo $C = C_s$. Además, se puede demostrar, de forma análoga al resultado del Teorema 4.3.2, que estos nuevos estadísticos se siguen distribuyendo como una combinación lineal de χ^2 con un grado de libertad.

4.4. Potencia Asintótica de los Estadísticos

En esta sección, se derivan las expresiones de la potencia asintótica de los estadísticos (4.2.19)–(4.2.22) para el caso de una covariable. No obstante el resultado se puede extender a varias covariables.

La *potencia exacta* de un contraste de hipótesis, con un nivel de significación α , basado en un estadístico T_n es la probabilidad π_n^E de rechazar H_0 cuando H_1 es verdadero y viene dada por

$$\pi_n^E(T_n) = P(T_n > \chi_1^2(1 - \alpha) \mid H_1),$$

donde $\chi_1^2(1 - \alpha)$ es el percentil de la distribución χ^2 con un grado de libertad.

Teorema 4.4.1. *Bajo las suposiciones del Corolario 4.3.1 y del Teorema 3.4.7, la potencia asintótica bajo $H_1 : \beta = \beta_1$, con un nivel de significación α , de los estadísticos $Kullback_1^{(1)}$, $Kullback_2^{(1)}$, $Jeffreys^{(1)}$ y $Renyi_\gamma^{(1)}$ viene dada por*

$$\pi_n(Kullback_1^{(1)}) = 1 - \Phi \left(\frac{\chi_1^2(1 - \alpha) \frac{\sigma_0^2 \sum_{i=1}^n [\varphi'(z_i, 0)]^2}{2C^2} - KL(F : F_0 | \beta_1)}{\sqrt{V[KL(F : F_0 | \hat{\beta}_n)]}} \right),$$

$$\pi_n(Kullback_2^{(1)}) = 1 - \Phi \left(\frac{\chi_1^2(1 - \alpha) \frac{\sigma_0^2 \sum_{i=1}^n [\varphi'(z_i, 0)]^2}{2C^2} - KL(F_0 : F | \beta_1)}{\sqrt{V[KL(F_0 : F | \hat{\beta}_n)]}} \right),$$

$$\pi_n(Jeffreys^{(1)}) = 1 - \Phi \left(\frac{\chi_1^2(1 - \alpha) \frac{\sigma_0^2 \sum_{i=1}^n [\varphi'(z_i, 0)]^2}{C^2} - J(F : F_0 | \beta_1)}{\sqrt{V[J(F : F_0 | \hat{\beta}_n)]}} \right),$$

y

$$\pi_n(Renyi_\gamma^{(1)}) = 1 - \Phi \left(\frac{\chi_1^2(1 - \alpha) \frac{\gamma \sigma_0^2 \sum_{i=1}^n [\varphi'(z_i, 0)]^2}{2C^2} - R_\gamma(F_0 : F | \beta_1)}{\sqrt{V[R_\gamma(F_0 : F | \hat{\beta}_n)]}} \right),$$

respectivamente, donde Φ es la función de distribución acumulativa de la distribución normal estándar y $V[T_n]$ denota la varianza de T_n .

Demostración. Para el estadístico $Kullback_1^{(1)}$ se rechaza H_0 , con un nivel de significación α , si

$$Kullback_1^{(1)}(\hat{\beta}_n) > \chi_1^2(1 - \alpha).$$

Entonces, la potencia exacta π_n^E del estadístico $Kullback_1^{(1)}$ para $H_1 : \beta = \beta_1$ es

$$\begin{aligned} \pi_n^E(Kullback_1^{(1)}) &= P \left(\frac{2C^2 KL(F : F_0 | \hat{\beta}_n)}{\sigma_0^2 \sum_{i=1}^n [\varphi'(z_i, 0)]^2} > \chi_1^2(1 - \alpha) \right) \\ &= P \left(KL(F : F_0 | \hat{\beta}_n) > \chi_1^2(1 - \alpha) \frac{\sigma_0^2 \sum_{i=1}^n [\varphi'(z_i, 0)]^2}{2C^2} \right) \\ &= P \left(\frac{KL(F : F_0 | \hat{\beta}_n) - KL(F : F_0 | \beta_1)}{\sqrt{V[KL(F : F_0 | \hat{\beta}_n)]}} \right. \\ &\quad \left. > \frac{\chi_1^2(1 - \alpha) \frac{\sigma_0^2 \sum_{i=1}^n [\varphi'(z_i, 0)]^2}{2C^2} - KL(F : F_0 | \beta_1)}{\sqrt{V[KL(F : F_0 | \hat{\beta}_n)]}} \right). \end{aligned}$$

Por el Corolario 4.3.1 y el Teorema 3.4.7, bajo la hipótesis alternativa H_1 se tiene $\sqrt{n}\hat{\beta}_n \sim N(\beta_1, \sigma_1^2)$, entonces $E[KL(F : F_0 | \hat{\beta}_n)] = KL(F : F_0 | \beta_1)$, por tanto

$$\frac{KL(F : F_0 | \hat{\beta}_n) - KL(F : F_0 | \beta_1)}{\sqrt{V[KL(F : F_0 | \hat{\beta}_n)]}} \sim N(0, 1)$$

cuando $n \rightarrow \infty$, de donde la potencia asintótica de $Kullback_1^{(1)}$ es

$$1 - \Phi \left(\frac{\chi_1^2(1 - \alpha) \frac{\sigma_0^2 \sum_{i=1}^n [\varphi'(z_i, 0)]^2}{2C^2} - KL(F : F_0 | \beta_1)}{\sqrt{V[KL(F : F_0 | \hat{\beta}_n)]}} \right).$$

La demostración para obtener la potencia asintótica de $Kullback_2^{(1)}$, $Jeffreys^{(1)}$ y $Renyi_\gamma^{(1)}$ es análoga. \square

Observación 4.4.2. Para obtener la varianza de los estadísticos $T_n(\hat{\beta}_n)$ bajo H_1 , se puede usar la aproximación de primer orden del desarrollo de Taylor, en torno a β_1 , esto es

$$\begin{aligned} T_n(\hat{\beta}_n) &\approx T_n(\beta_1) + (\hat{\beta}_n - \beta_1) \frac{\partial}{\partial \beta} T_n(\beta_1) \\ &= T_n(\beta_1) + \hat{\beta}_n \frac{\partial}{\partial \beta} T_n(\beta_1) - \beta_1 \frac{\partial}{\partial \beta} T_n(\beta_1) \end{aligned}$$

así

$$Var[T_n(\hat{\beta}_n)] \approx \left[\frac{\partial}{\partial \beta} T_n(\beta_1) \right]^2 \sigma_1^2,$$

donde las derivadas de los estadísticos evaluadas en β_1 vienen dadas por

$$\frac{\partial}{\partial \beta} KL(F : F_0 | \beta_1) = \sum_{i=1}^n \left(\frac{\varphi'(z_i, \beta_1)}{\varphi(z_i, \beta_1)} - C \frac{\varphi'(z_i, \beta_1)}{[\varphi(z_i, \beta_1)]^2} \right),$$

$$\frac{\partial}{\partial \beta} KL(F_0 : F | \beta_1) = \sum_{i=1}^n \left(-\frac{\varphi'(z_i, \beta_1)}{\varphi(z_i, \beta_1)} + \frac{\varphi'(z_i, \beta_1)}{C} \right),$$

$$\frac{\partial}{\partial \beta} J(F : F_0 | \beta_1) = \sum_{i=1}^n \left(\frac{\varphi'(z_i, \beta_1)}{\varphi(z_i, \beta_1)} - C \frac{\varphi'(z_i, \beta_1)}{[\varphi(z_i, \beta_1)]^2} \right),$$

y

$$\frac{\partial}{\partial \beta} R_\gamma(F_0 : F | \beta_1) = 2n \frac{1}{1-\gamma} \sum_{i=1}^n \frac{(1-\gamma)\varphi'(z_i, \beta_1)}{(1-\gamma)\varphi(z_i, \beta_1) + \gamma C} - \sum_{i=1}^n \frac{\varphi'(z_i, \beta_1)}{\varphi(z_i, \beta_1)},$$

respectivamente.

4.5. Comparación de los Estadísticos Propuestos

En esta sección, se comparan los nuevos estadísticos definidos anteriormente con los estadísticos clásicos para el modelo de Cox (4.1.1) y el modelo generalizado

de riesgos proporcionales (4.1.2), para datos doblemente censurados y datos con censura por la derecha, respectivamente. Para ello se lleva a cabo el estudio de simulación que se describe a continuación.

De acuerdo a las Secciones 1.2.3 y 2.4, los tiempos de supervivencia t se generan bajo el modelo (4.1.4) mediante la expresión

$$t = \Lambda_0^{-1} \left(-\frac{\log u}{\varphi(\mathbf{z}, \boldsymbol{\beta})} \right),$$

donde $u \sim \mathcal{U}(0, 1)$ y $\varphi(\mathbf{z}, \boldsymbol{\beta}) = \exp(\boldsymbol{\beta}'\mathbf{z})$ y $\varphi(\mathbf{z}, \boldsymbol{\beta}) = K(\boldsymbol{\beta}, \mathbf{z}, a)$ como en (2.2.2) con $a > 0$ para el modelo (4.1.1) y (4.1.2), respectivamente. Se asume que la función de riesgo base acumulativa $\Lambda_0(\cdot)$ es una función Weibull con parámetros $\nu = 3.2$ y $\gamma = 1.1$.

Se estudia y analiza el caso 1-dimensional, tal que $\boldsymbol{\beta}'\mathbf{Z} = \beta Z$, con $Z \sim \mathcal{U}(0, 1)$. Se toman tamaños de muestra $n = 20, 50, 100$ y 200 para 5000 bases de datos simuladas.

Se comparan los estadísticos $Kullback_1^{(1)}$, $Kullback_2^{(1)}$, $Jeffreys^{(1)}$ y $Renyi^{(1)}$ con $\gamma = 0.10, 0.25, 0.50$ y 0.75 , además del estadístico de *Wald* y *LR*. Para llevar a cabo un estudio comparativo de los estadísticos anteriores, denotados por T_n , se calcula tanto los tamaños exactos (los datos se simulan bajo $H_0 : \beta = 0$), como las potencias exactas (los datos se simulan bajo $H_1 : \beta = 1.5$), para una muestra de tamaño n . Estos vienen dados por

$$\alpha_n^E(T_n) = P(T_n > \chi_1^2(1 - \alpha) \mid H_0)$$

y

$$\pi_n^E(T_n) = P(T_n > \chi_1^2(1 - \alpha) \mid H_1),$$

respectivamente. El cálculo, tanto del tamaño exacto ($H_\delta, \delta = 0$), como de la potencia exacta ($H_\delta, \delta = 1$) se lleva a cabo mediante simulación utilizando la fórmula

$$\frac{\text{Número de } (T_{nj} > \chi_1^2(1 - \alpha) \mid H_\delta)}{N}$$

donde T_{nj} , $j = 1, \dots, N$ es el estadístico T_n para la j -ésima muestra simulada con $N = 5000$ el número de simulaciones realizadas y $\alpha = 0.05$ ó 0.10 .

En las Tablas 4.1, 4.2, 4.3 y 4.4 se muestran los resultados para datos provenientes de un modelo de Cox doblemente censurados. En las Tablas 4.1 y 4.2 se muestran los resultados de tamaño exacto (primera columna), potencia exacta (segunda columna), potencia asintótica (tercera columna) y la diferencia entre potencia exacta y tamaño exacto (cuarta columna) para tamaños nominales de $\alpha = 0.05$ y $\alpha = 0.10$, respectivamente, para un conjunto de datos tal que el 50 % de datos son completos, 45 % de datos son con censura por la derecha y 5 % de datos son con censura por izquierda, aproximadamente. De forma análoga en las Tablas 4.3 y 4.4 se muestran los resultados para un conjunto de datos tal que el 50 % de datos son completos, 20 % de datos son con censura por la derecha y 30 % de datos son con censura por izquierda, aproximadamente.

En las Tablas 4.1, 4.2, 4.3 y 4.4, para cada tamaño de muestra n se han señalado en bold los 3 estadísticos con mejor tamaño exacto en cada caso y se observa que el tamaño exacto del estadístico de $Renyi_{0,50}$ es más cercano al tamaño nominal que el resto, incluso que los estadísticos clásicos. Sin embargo, de los 3 estadísticos marcados, los estadísticos $Renyi_{0,50}$, $Renyi_{0,75}$ o el de $Wald$, tienen mayor potencia exacta dependiendo del caso (se ha marcado en bold). Se observa en la última columna de las tablas que la diferencia entre potencia y tamaños exactos es mayor para $Renyi_{0,75}$ en más de la mitad de los casos. El mayor valor de entre los tres estadísticos marcados en bold es marcado en bold en esta columna. Por tanto, el estadístico de $Renyi_{0,75}$ es una buena alternativa para llevar a cabo este contraste y mejor que los estadísticos clásicos, $Wald$ y LR . Además, se observa que cuando n crece la potencia asintótica de los estadísticos se aproxima a la exacta. Claramente para tamaños de muestra pequeños la potencia asintótica sobreestima la exacta. Una forma rápida de observar el comportamiento de estos estimadores es graficar

los valores obtenidos en las Tablas 4.1–4.4. En las Figuras 4.1 y 4.6 se grafican los valores Tamaño exacto - Tamaño nominal vs. $\log(n)$ y Potencia exacta - Potencia asintótica vs. $\log(n)$ para ambas proporciones de censura, respectivamente y para tamaños nominales de 0.05 y 0.10. De las Figuras 4.1 A. y 4.6 A., se puede concluir que el estadístico $Renyi_{0,5}$ es el que tiene mejor tamaño exacto. Sin embargo, de las Figuras 4.1 B. y 4.6 B. se observa que, en cuanto Potencia exacta - Potencia asintótica el estadístico LR no es una buena opción para n pequeño, pero es el mejor cuando n crece.

En las Figuras 4.2, 4.3, 4.4, 4.7, 4.8 y 4.9 se muestra la función de densidad estimada de los valores de los estadísticos para ambas configuraciones de proporciones de censura y la función de densidad de una χ_1^2 , para $n = 50$ y 100 . Se puede observar que la función de densidad estimada, en todos los casos, es cercana a una función de distribución χ_1^2 . No obstante, puesto que para llevar a cabo el contraste de hipótesis lo más importante es el comportamiento de la cola, en las Figuras 4.5 y 4.10 se representan las colas de las densidades para los 3 estadísticos mejores: $Renyi_{0,5}$, $Renyi_{0,75}$ y $Wald$. De estas gráficas se puede concluir que la distribución exacta simulada del estadístico de $Renyi_{0,5}$ es la que más se aproxima a la distribución asintótica que es la que se utiliza para la construcción de la región crítica del contraste.

En las Tablas 4.5 y 4.6 se muestran los resultados para datos provenientes de un modelo generalizado de riesgos proporcionales estudiado en el Capítulo 2 con 25 % de censura por la derecha. Se muestran el tamaño exacto, potencia exacta y potencia asintótica, así como la diferencia entre tamaño y potencia exacta. Al igual que antes, para cada tamaño de muestra se han marcado los 3 estadísticos con mejor tamaño exacto. Se puede observar que los estadístico de $Kullback_1$ y $Renyi_{0,1}$ tienen buenos resultados en cuanto a tamaño, sin embargo el estadístico LR es mejor en la mayoría de los casos. Por otro lado, de los 3 estadísticos marcados, en cuanto

Inferencia para los Modelos

n	Estadístico	Tamaño exacto	Potencia exacta	Potencia asintótica	PE - TE
20	$Kullback_1$	0.1012	0.0998	0.2149	-0.0014
	$Kullback_2$	0.1018	0.4324	0.7057	0.3306
	$Jeffreys$	0.1254	0.3410	0.5670	0.2156
	$Renyi_{0,10}$	0.0874	0.1168	0.2505	0.0294
	$Renyi_{0,25}$	0.0702	0.1452	0.3112	0.0750
	$Renyi_{0,50}$	0.0614	0.2142	0.4306	0.1528
	$Renyi_{0,75}$	0.0698	0.3116	0.5667	0.2418
	$Wald$	0.0842	0.2614	0.5437	0.1772
	LR	0.0738	0.2496	0.5394	0.1758
50	$Kullback_1$	0.0690	0.3878	0.7081	0.3188
	$Kullback_2$	0.0710	0.6830	0.8980	0.6120
	$Jeffreys$	0.0756	0.6050	0.8617	0.5294
	$Renyi_{0,10}$	0.0620	0.4230	0.7347	0.3610
	$Renyi_{0,25}$	0.0570	0.4682	0.7720	0.4112
	$Renyi_{0,50}$	0.0554	0.5472	0.8260	0.4918
	$Renyi_{0,75}$	0.0602	0.6174	0.8686	0.5572
	$Wald$	0.0618	0.5666	0.8413	0.5048
	LR	0.0612	0.5738	0.6913	0.5126
100	$Kullback_1$	0.0594	0.8042	0.9420	0.7448
	$Kullback_2$	0.0662	0.9122	0.9650	0.8460
	$Jeffreys$	0.0666	0.8820	0.9602	0.8154
	$Renyi_{0,10}$	0.0592	0.8170	0.9468	0.7578
	$Renyi_{0,25}$	0.0568	0.8374	0.9530	0.7806
	$Renyi_{0,50}$	0.0546	0.8680	0.9608	0.8134
	$Renyi_{0,75}$	0.0592	0.8898	0.9654	0.8306
	$Wald$	0.0586	0.8720	0.9592	0.8134
	LR	0.0612	0.8768	0.8679	0.8156
200	$Kullback_1$	0.0590	0.9914	0.9960	0.9324
	$Kullback_2$	0.0596	0.9964	0.9926	0.9368
	$Jeffreys$	0.0578	0.9948	0.9937	0.9370
	$Renyi_{0,10}$	0.0546	0.9916	0.9961	0.9370
	$Renyi_{0,25}$	0.0522	0.9928	0.9962	0.9406
	$Renyi_{0,50}$	0.0530	0.9944	0.9961	0.9414
	$Renyi_{0,75}$	0.0552	0.9952	0.9953	0.9400
	$Wald$	0.0540	0.9944	0.9951	0.9404
	LR	0.0550	0.9946	0.9729	0.9396

Tabla 4.1: Tamaño exacto, Potencia exacta, Potencia asintótica y Potencia exacta - Tamaño exacto para bases de datos con aproximadamente 50 %, 45 % y 5 % de datos completos, censurados por la derecha e izquierda, respectivamente, ajustados al modelo de Cox con $\alpha = 0.050$.

Inferencia para los Modelos

n	Estadístico	Tamaño exacto	Potencia exacta	Potencia asintótica	PE - TE
20	$Kullback_1$	0.1498	0.2030	0.4297	0.0532
	$Kullback_2$	0.1498	0.5002	0.7894	0.3504
	$Jeffreys$	0.1802	0.4330	0.7047	0.2528
	$Renyi_{0,10}$	0.1384	0.2262	0.4655	0.0878
	$Renyi_{0,25}$	0.1246	0.2620	0.5210	0.1374
	$Renyi_{0,50}$	0.1144	0.3396	0.6154	0.2252
	$Renyi_{0,75}$	0.1214	0.4202	0.7072	0.2988
	$Wald$	0.1370	0.3758	0.6906	0.2388
	LR	0.1296	0.3696	0.6582	0.2400
50	$Kullback_1$	0.1136	0.5762	0.8431	0.4626
	$Kullback_2$	0.1256	0.7630	0.9232	0.6374
	$Jeffreys$	0.1326	0.7050	0.9071	0.5724
	$Renyi_{0,10}$	0.1098	0.5972	0.8554	0.4874
	$Renyi_{0,25}$	0.1104	0.6270	0.8722	0.5166
	$Renyi_{0,50}$	0.1114	0.6768	0.8957	0.5654
	$Renyi_{0,75}$	0.1162	0.7176	0.9132	0.6014
	$Wald$	0.1198	0.6858	0.8981	0.5660
	LR	0.1188	0.6906	0.7750	0.5718
100	$Kullback_1$	0.1134	0.8968	0.9713	0.7834
	$Kullback_2$	0.1138	0.9454	0.9725	0.8316
	$Jeffreys$	0.1170	0.9284	0.9723	0.8114
	$Renyi_{0,10}$	0.1122	0.9034	0.9727	0.7912
	$Renyi_{0,25}$	0.1112	0.9114	0.9744	0.8002
	$Renyi_{0,50}$	0.1084	0.9244	0.9763	0.8160
	$Renyi_{0,75}$	0.1102	0.9354	0.9763	0.8252
	$Wald$	0.1126	0.9256	0.9735	0.8130
	LR	0.1138	0.9278	0.9028	0.8140
200	$Kullback_1$	0.1072	0.9966	0.9980	0.8894
	$Kullback_2$	0.1104	0.9986	0.9940	0.8882
	$Jeffreys$	0.1106	0.9982	0.9954	0.8876
	$Renyi_{0,10}$	0.1078	0.9970	0.9980	0.8892
	$Renyi_{0,25}$	0.1070	0.9976	0.9979	0.8906
	$Renyi_{0,50}$	0.1070	0.9982	0.9975	0.8912
	$Renyi_{0,75}$	0.1072	0.9984	0.9967	0.8912
	$Wald$	0.1080	0.9982	0.9967	0.8902
	LR	0.1096	0.9982	0.9819	0.8886

Tabla 4.2: Tamaño exacto, Potencia exacta, Potencia asintótica y Potencia exacta - Tamaño exacto para bases de datos con aproximadamente 50 %, 45 % y 5 % de datos completos, censurados por la derecha e izquierda, respectivamente, ajustados al modelo de Cox con $\alpha = 0.100$.

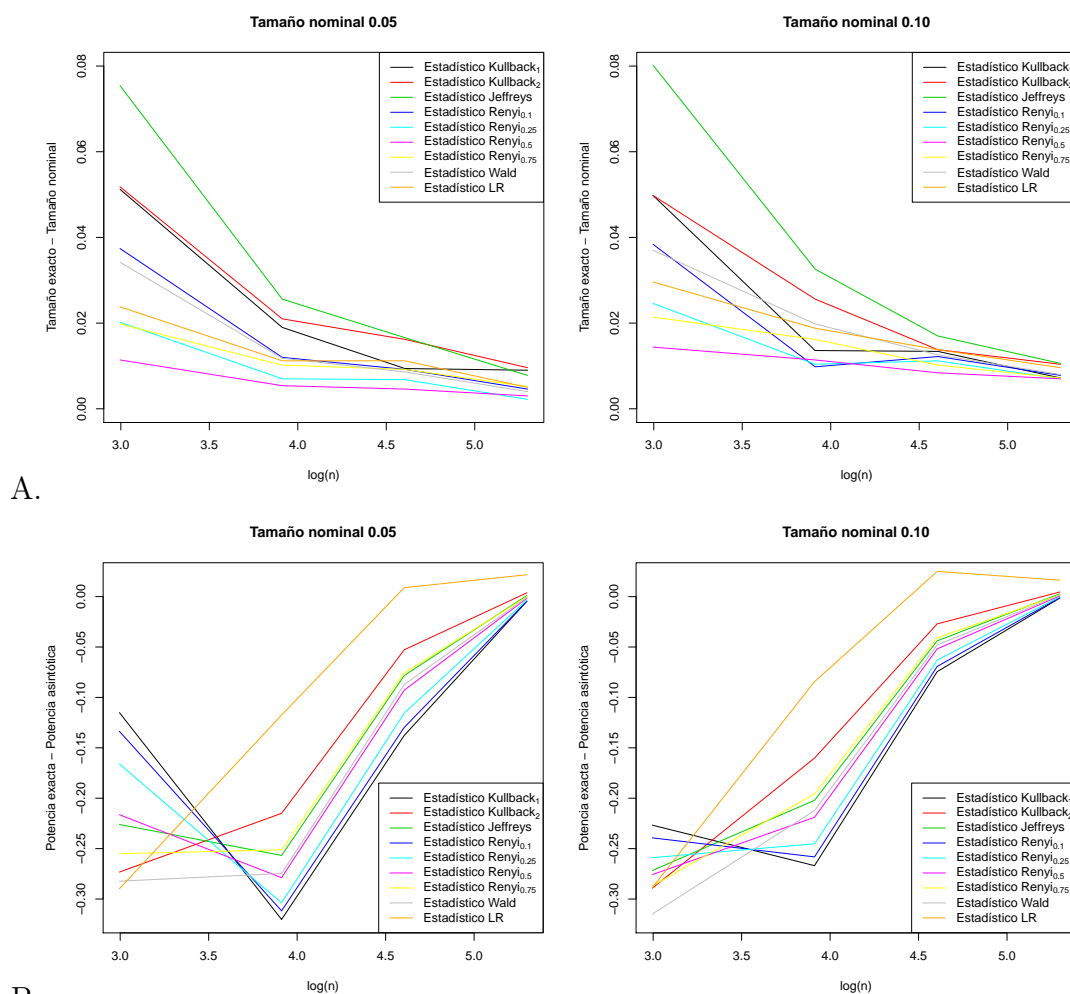


Figura 4.1: A. Gráfica tamaño exacto - tamaño nominal vs. $\log(n)$ y B. Gráfica potencia exacta - potencia asintótica vs. $\log(n)$ para el modelo de Cox con aproximadamente 50 %, 45 % y 5 % de datos completos, censurados por la derecha e izquierda, respectivamente.

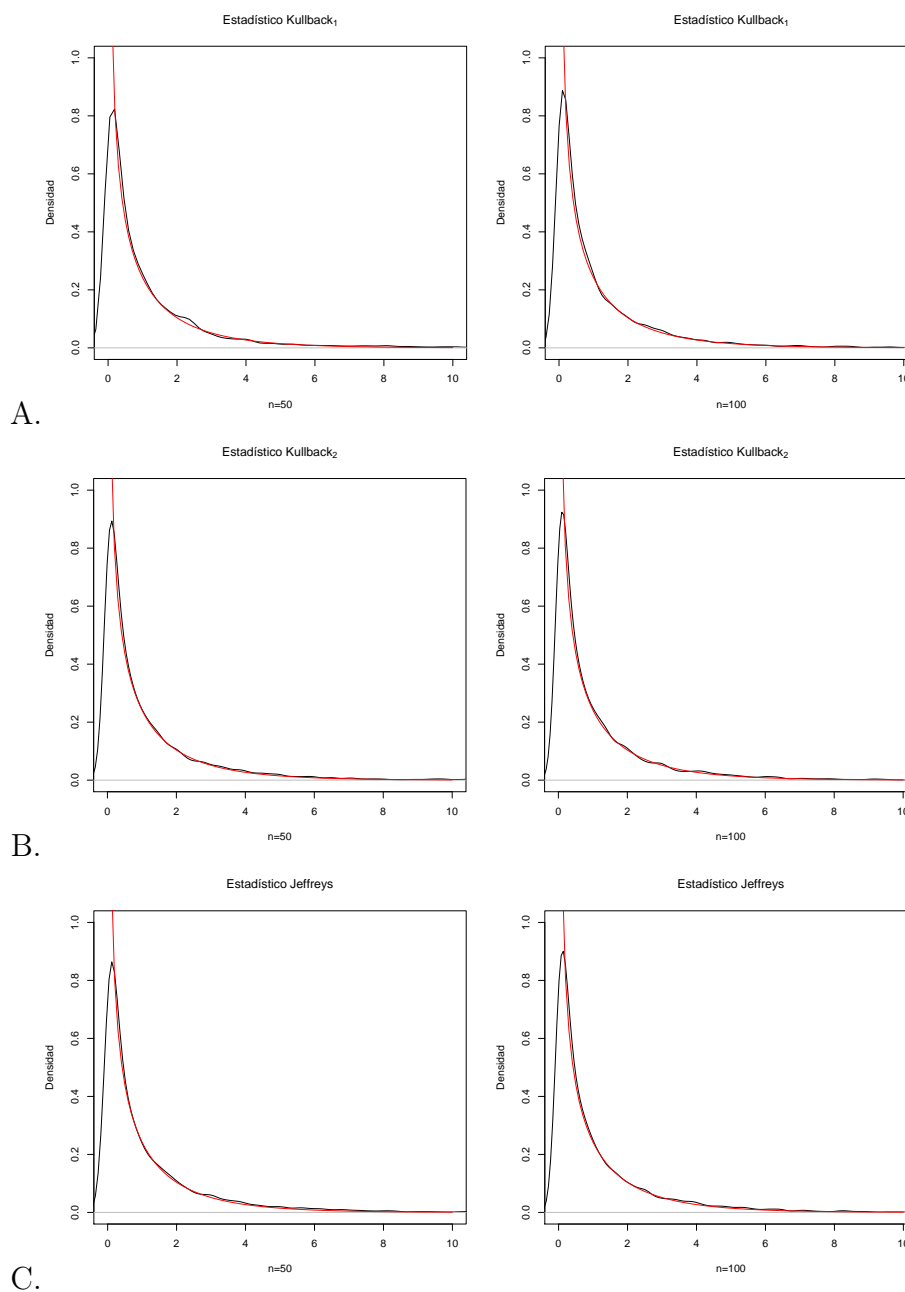


Figura 4.2: Gráfica de la densidad empírica de los estimadores de los estadísticos A. $Kullback_1$, B. $Kullback_2$ y C. $Jeffreys$ bajo H_0 con $n=50$ y $n=100$ para el modelo de Cox con aproximadamente 50 %, 45 % y 5 % de datos completos, censurados por la derecha e izquierda, respectivamente, comparada con la gráfica (en rojo) de la densidad χ^2_1 .

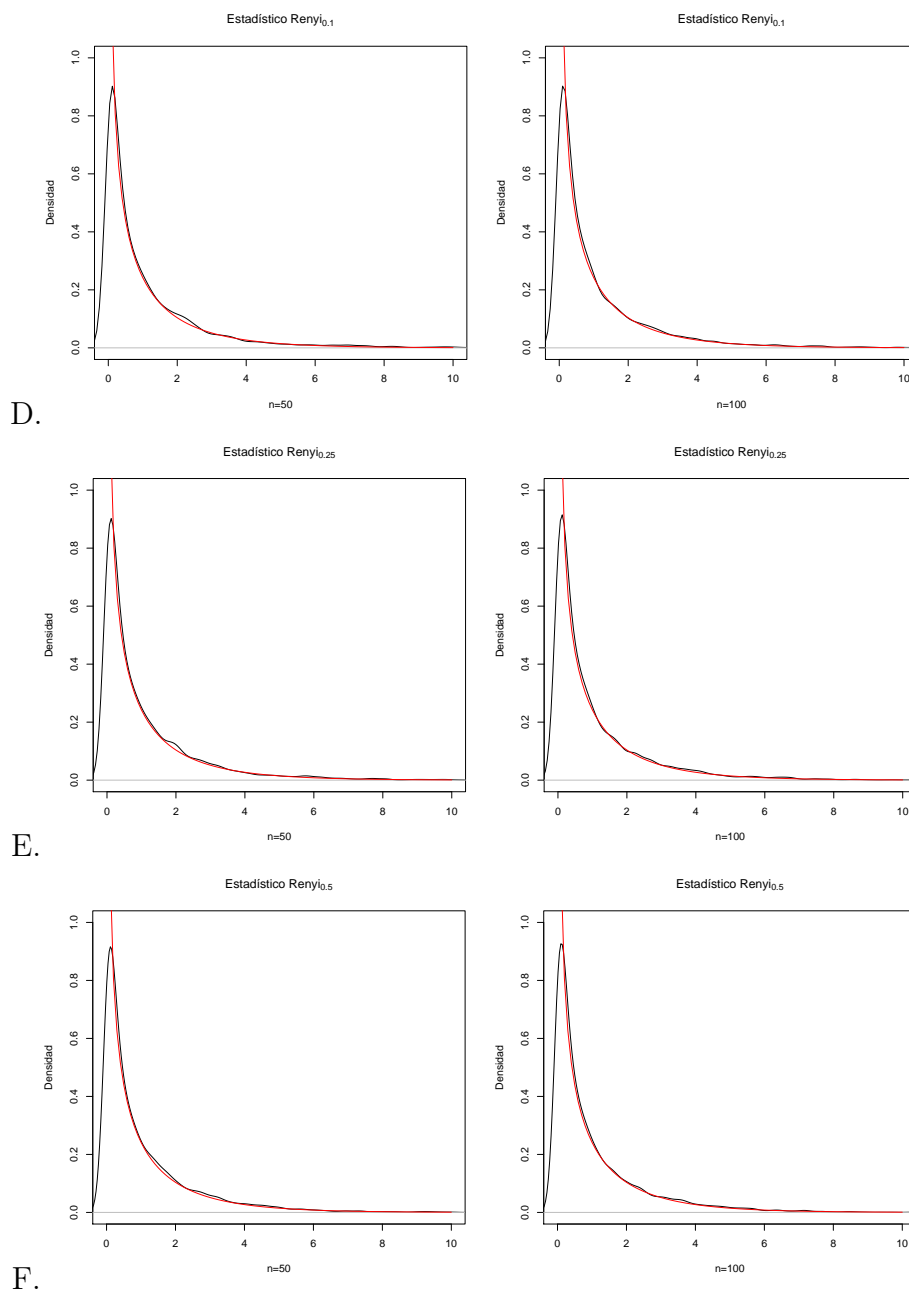


Figura 4.3: Gráfica de la densidad empírica de los estimadores de los estadísticos D. $Renyi_{0,10}$, E. $Renyi_{0,25}$ y F. $Renyi_{0,50}$ bajo H_0 con $n=50$ y $n=100$ para el modelo de Cox con aproximadamente 50 %, 45 % y 5 % de datos completos, censurados por la derecha e izquierda, respectivamente, comparada con la gráfica (en rojo) de la densidad χ^2_1 .

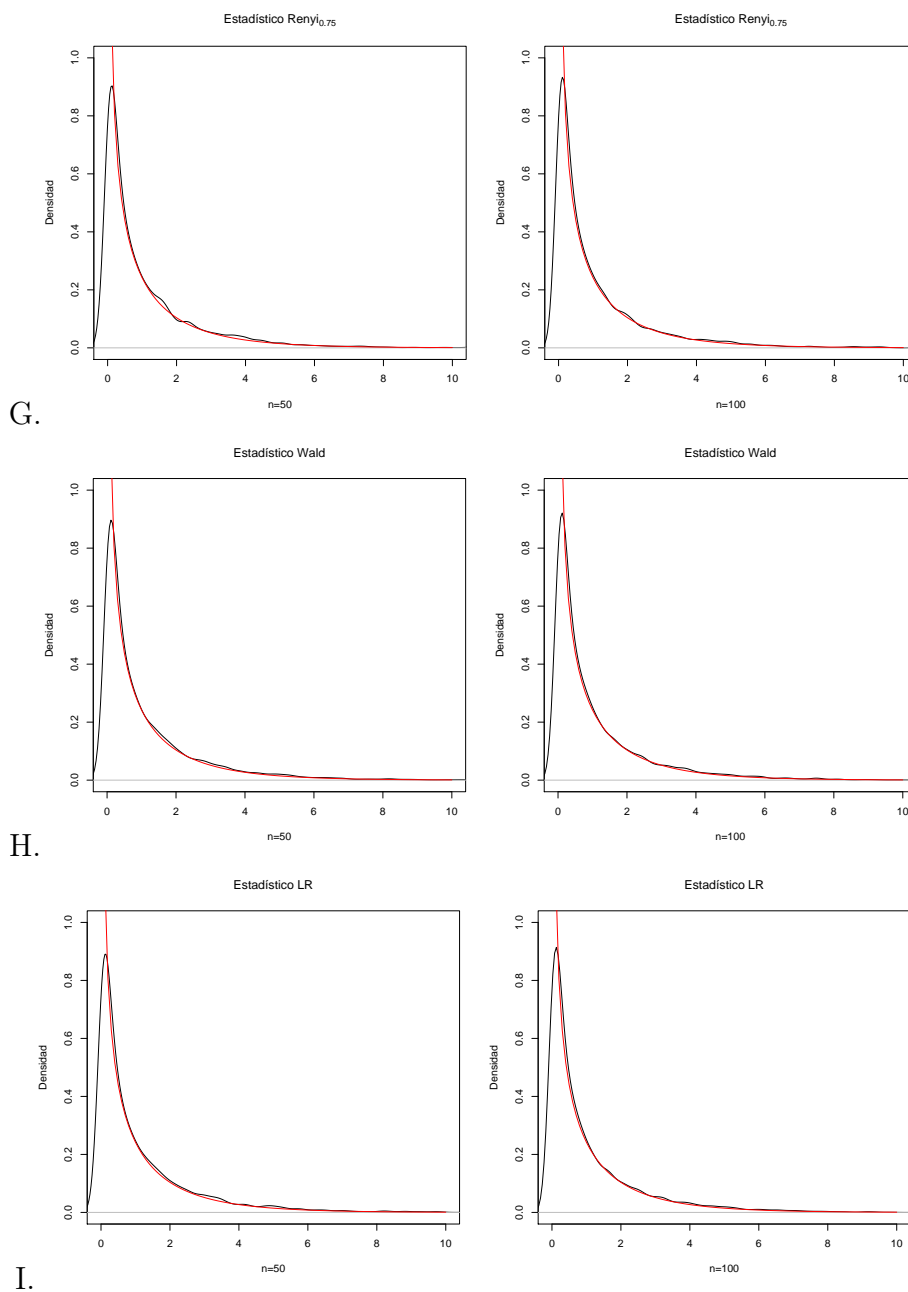


Figura 4.4: Gráfica de la densidad empírica de los estimadores de los estadísticos G. $Renyi_{0.75}$, H. $Wald$ e I. LR bajo H_0 con $n=50$ y $n=100$ para el modelo de Cox con aproximadamente 50 %, 45 % y 5 % de datos completos, censurados por la derecha e izquierda, respectivamente, comparada con la gráfica (en rojo) de la densidad χ^2_1 .

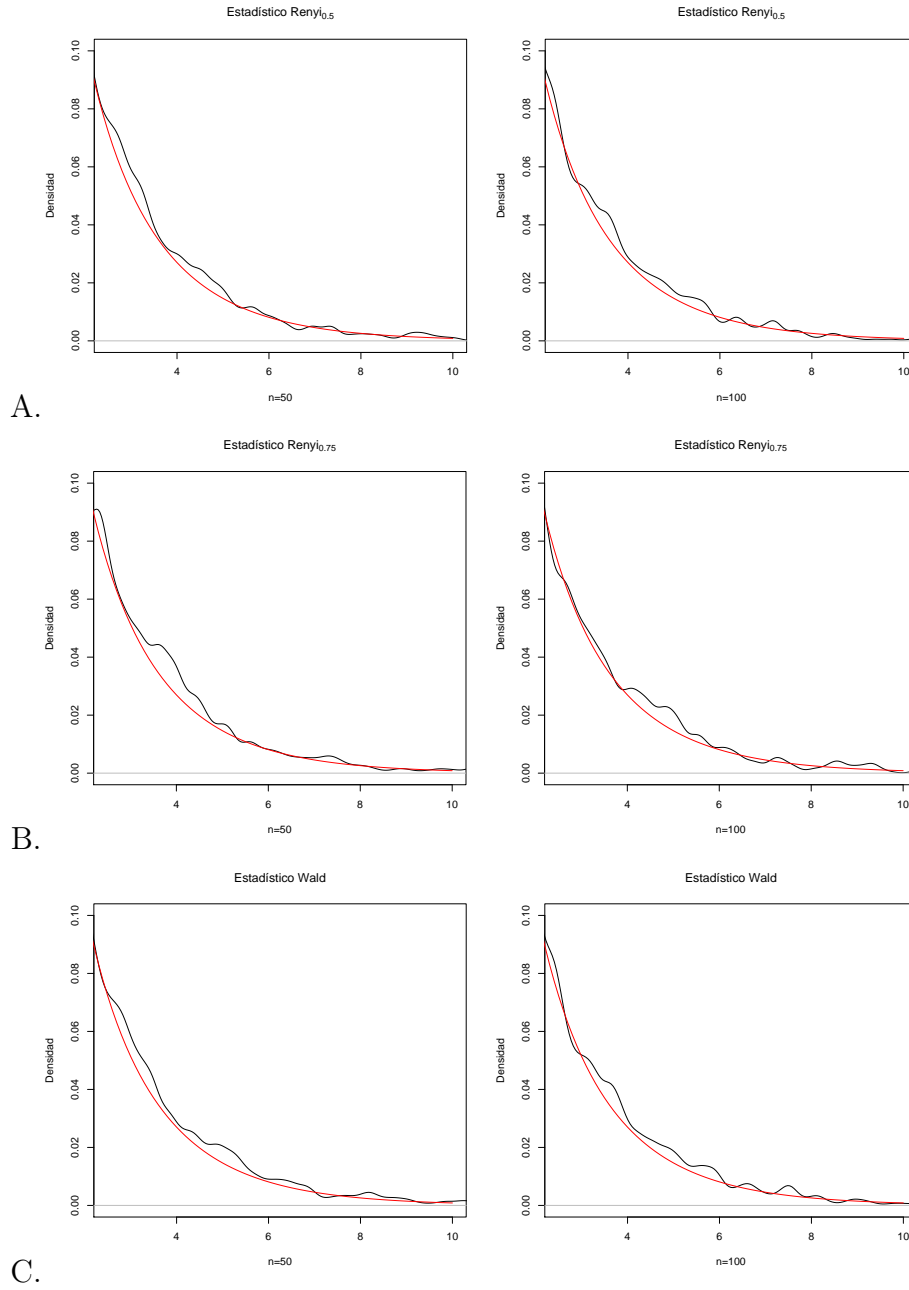


Figura 4.5: Gráfica de la densidad empírica de los estimadores de los estadísticos en las colas A. $Renyi_{0,50}$, B. $Renyi_{0,75}$ y C. $Wald$ bajo H_0 con $n=50$ y $n=100$ para el modelo de Cox con aproximadamente 50 %, 45 % y 5 % de datos completos, censurados por la derecha e izquierda, respectivamente, comparada con la gráfica (en rojo) de la densidad χ^2_1 .

Inferencia para los Modelos

n	Estadístico	Tamaño exacto	Potencia exacta	Potencia asintótica	PE - TE
20	$Kullback_1$	0.1138	0.1192	0.2354	0.0054
	$Kullback_2$	0.1112	0.4648	0.7177	0.3536
	$Jeffreys$	0.1390	0.3850	0.5843	0.2460
	$Renyi_{0,10}$	0.1000	0.1394	0.2719	0.0394
	$Renyi_{0,25}$	0.0818	0.1756	0.3334	0.0938
	$Renyi_{0,50}$	0.0698	0.2612	0.4521	0.1914
	$Renyi_{0,75}$	0.0812	0.3598	0.5846	0.2786
	$Wald$	0.0912	0.3072	0.5578	0.2160
	LR	0.0854	0.2992	0.5530	0.2138
50	$Kullback_1$	0.0786	0.4456	0.7332	0.3670
	$Kullback_2$	0.0804	0.7274	0.9048	0.6470
	$Jeffreys$	0.0842	0.6532	0.8721	0.5690
	$Renyi_{0,10}$	0.0744	0.4804	0.7578	0.4060
	$Renyi_{0,25}$	0.0676	0.5230	0.7920	0.4554
	$Renyi_{0,50}$	0.0638	0.5958	0.8410	0.5320
	$Renyi_{0,75}$	0.0706	0.6662	0.8791	0.5956
	$Wald$	0.0702	0.6198	0.8536	0.5496
	LR	0.0720	0.6240	0.7281	0.5520
100	$Kullback_1$	0.0646	0.8460	0.9493	0.7814
	$Kullback_2$	0.0622	0.9310	0.9678	0.8688
	$Jeffreys$	0.0640	0.9082	0.9640	0.8442
	$Renyi_{0,10}$	0.0602	0.8584	0.9534	0.7982
	$Renyi_{0,25}$	0.0588	0.8732	0.9587	0.8144
	$Renyi_{0,50}$	0.0588	0.8982	0.9653	0.8418
	$Renyi_{0,75}$	0.0576	0.9144	0.9689	0.8568
	$Wald$	0.0594	0.9014	0.9635	0.8420
	LR	0.0620	0.9042	0.8925	0.8422
200	$Kullback_1$	0.0596	0.9932	0.9968	0.9336
	$Kullback_2$	0.0634	0.9972	0.9935	0.9338
	$Jeffreys$	0.0622	0.9960	0.9945	0.9338
	$Renyi_{0,10}$	0.0572	0.9938	0.9969	0.9366
	$Renyi_{0,25}$	0.0562	0.9944	0.9969	0.9382
	$Renyi_{0,50}$	0.0562	0.9958	0.9967	0.9396
	$Renyi_{0,75}$	0.0594	0.9960	0.9960	0.9366
	$Wald$	0.0582	0.9958	0.9959	0.9376
	LR	0.0594	0.9960	0.9756	0.9366

Tabla 4.3: Tamaño exacto, Potencia exacta, Potencia asintótica y Potencia exacta - Tamaño exacto para bases de datos con aproximadamente 50 %, 20 % y 30 % de datos completos, censurados por la derecha e izquierda, respectivamente, ajustados al modelo de Cox con $\alpha = 0.050$.

Inferencia para los Modelos

n	Estadístico	Tamaño exacto	Potencia exacta	Potencia asintótica	PE - TE
20	$Kullback_1$	0.1648	0.2478	0.4529	0.0830
	$Kullback_2$	0.1630	0.5388	0.7981	0.3758
	$Jeffreys$	0.1990	0.4676	0.7177	0.2686
	$Renyi_{0,10}$	0.1542	0.2736	0.4880	0.1194
	$Renyi_{0,25}$	0.1418	0.3100	0.5420	0.1682
	$Renyi_{0,50}$	0.1314	0.3822	0.6331	0.2508
	$Renyi_{0,75}$	0.1416	0.4566	0.7206	0.3150
	$Wald$	0.1578	0.4150	0.7014	0.2572
	LR	0.1512	0.4134	0.6674	0.2622
50	$Kullback_1$	0.1334	0.6326	0.8580	0.4992
	$Kullback_2$	0.1354	0.8028	0.9281	0.6674
	$Jeffreys$	0.1504	0.7504	0.9140	0.6000
	$Renyi_{0,10}$	0.1316	0.6510	0.8691	0.5194
	$Renyi_{0,25}$	0.1282	0.6778	0.8841	0.5496
	$Renyi_{0,50}$	0.1248	0.7250	0.9049	0.6002
	$Renyi_{0,75}$	0.1290	0.7646	0.9201	0.6356
	$Wald$	0.1330	0.7352	0.9062	0.6022
	LR	0.1384	0.7390	0.8051	0.6006
100	$Kullback_1$	0.1168	0.9216	0.9750	0.8048
	$Kullback_2$	0.1168	0.9590	0.9747	0.8422
	$Jeffreys$	0.1218	0.9466	0.9749	0.8248
	$Renyi_{0,10}$	0.1154	0.9260	0.9762	0.8106
	$Renyi_{0,25}$	0.1138	0.9316	0.9776	0.8178
	$Renyi_{0,50}$	0.1138	0.9428	0.9790	0.8300
	$Renyi_{0,75}$	0.1128	0.9494	0.9787	0.8366
	$Wald$	0.1158	0.9442	0.9764	0.8284
	LR	0.1182	0.9454	0.9205	0.8272
200	$Kullback_1$	0.1172	0.9972	0.9984	0.8800
	$Kullback_2$	0.1148	0.9980	0.9947	0.8832
	$Jeffreys$	0.1172	0.9978	0.9960	0.8806
	$Renyi_{0,10}$	0.1172	0.9974	0.9984	0.8802
	$Renyi_{0,25}$	0.1142	0.9976	0.9983	0.8834
	$Renyi_{0,50}$	0.1124	0.9978	0.9980	0.8854
	$Renyi_{0,75}$	0.1126	0.9978	0.9972	0.8852
	$Wald$	0.1136	0.9978	0.9972	0.8842
	LR	0.1158	0.9978	0.9827	0.8820

Tabla 4.4: Tamaño exacto, Potencia exacta, Potencia asintótica y Potencia exacta - Tamaño exacto para bases de datos con aproximadamente 50 %, 20 % y 30 % de datos completos, censurados por la derecha e izquierda, respectivamente, ajustados al modelo de Cox con $\alpha = 0.100$.

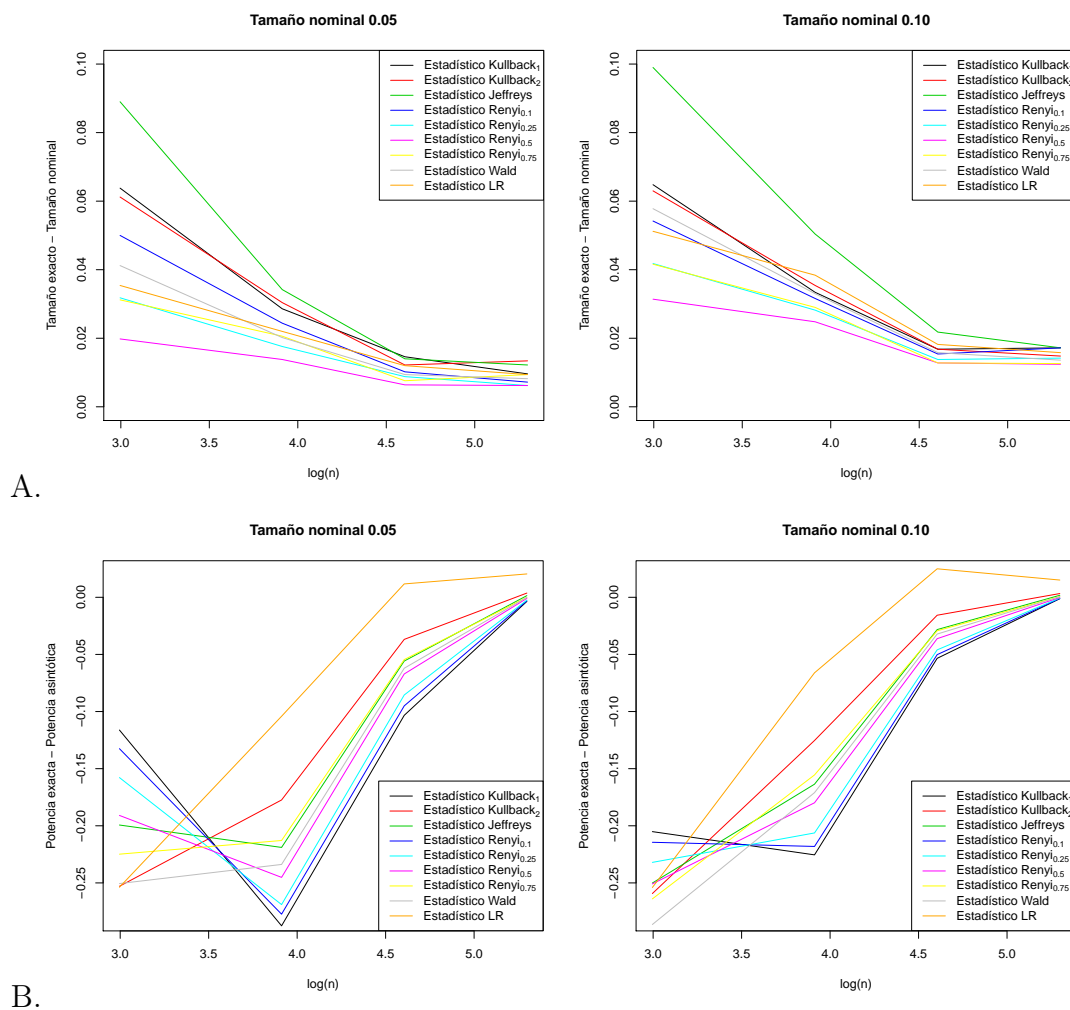


Figura 4.6: A. Gráfica tamaño exacto - tamaño nominal vs. $\log(n)$ y B. Gráfica potencia exacta - potencia asintótica vs. $\log(n)$ para el modelo de Cox con aproximadamente 50 %, 20 % y 30 % de datos completos, censurados por la derecha e izquierda, respectivamente.

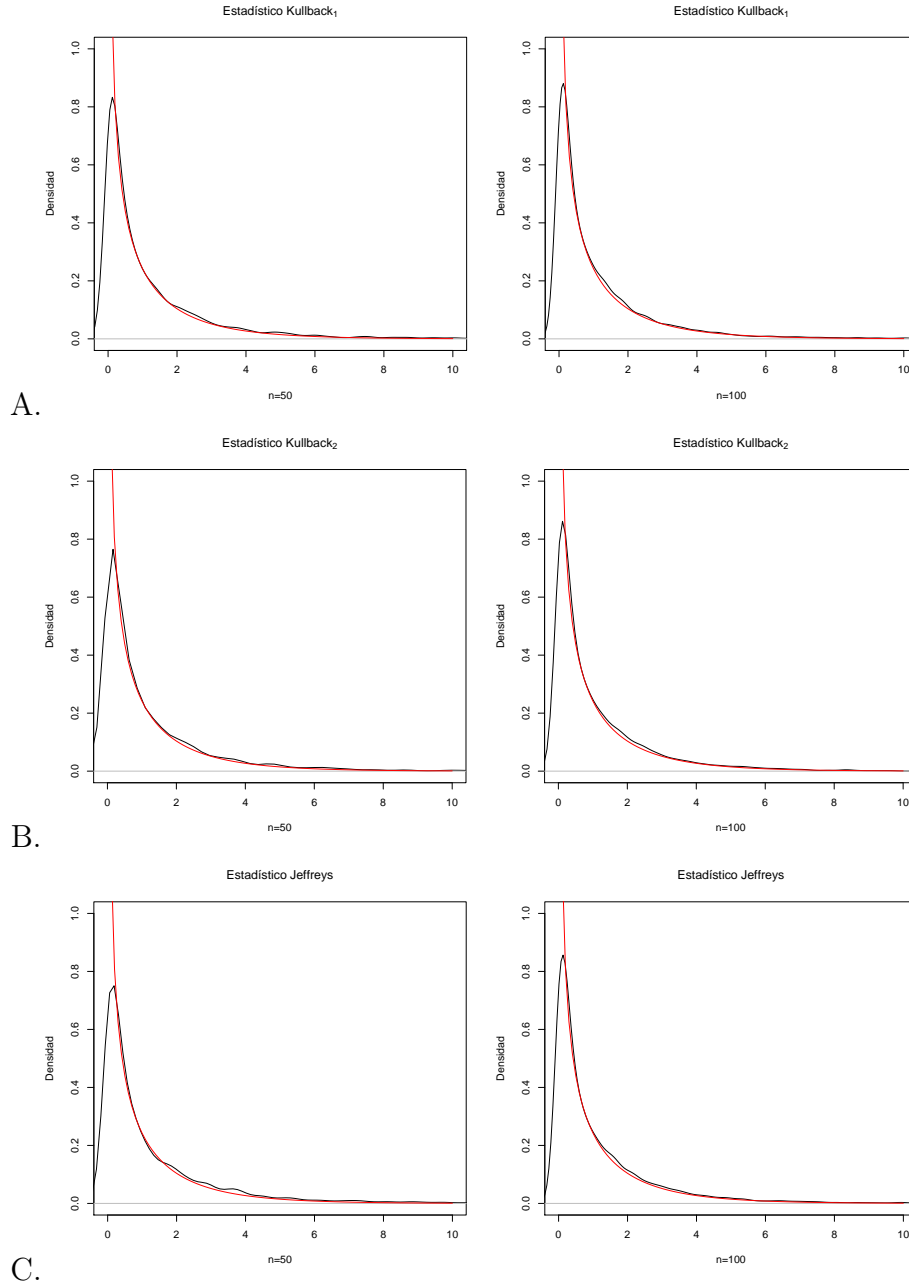


Figura 4.7: Gráfica de la densidad empírica de los estimadores de los estadísticos A. *Kullback*₁, B. *Kullback*₂ y C. *Jeffreys* bajo H_0 con $n=50$ y $n=100$ para el modelo de Cox con aproximadamente 50 %, 20 % y 30 % de datos completos, censurados por la derecha e izquierda, respectivamente, comparada con la gráfica (en rojo) de la densidad χ^2_1 .

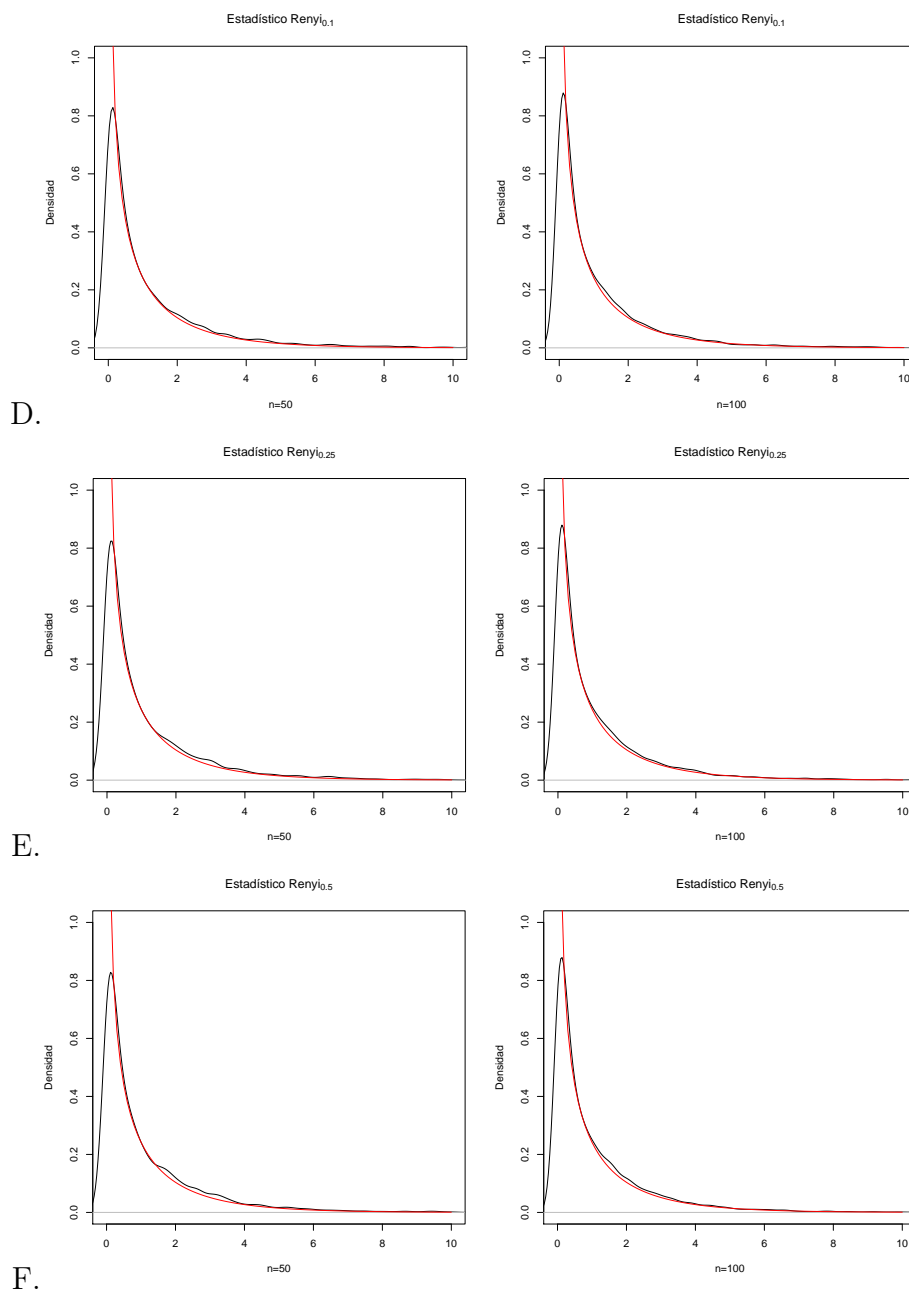


Figura 4.8: Gráfica de la densidad empírica de los estimadores de los estadísticos D. $Renyi_{0,10}$, E. $Renyi_{0,25}$ y F. $Renyi_{0,50}$ bajo H_0 con $n=50$ y $n=100$ para el modelo de Cox con aproximadamente 50 %, 20 % y 30 % de datos completos, censurados por la derecha e izquierda, respectivamente, comparada con la gráfica (en rojo) de la densidad χ^2_1 .

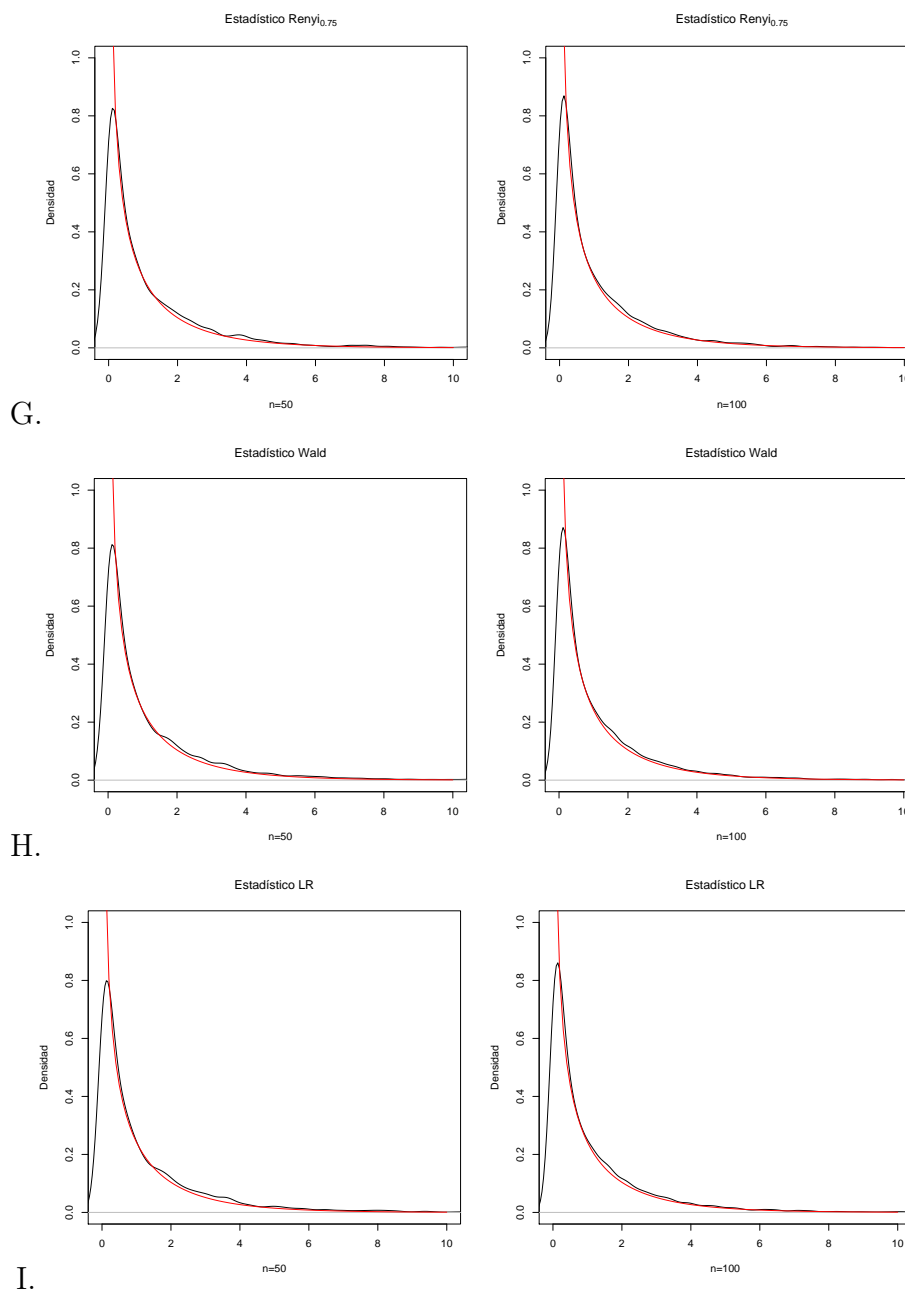


Figura 4.9: Gráfica de la densidad empírica de los estimadores de los estadísticos G. $Renyi_{0.75}$, H. $Wald$ e I. LR bajo H_0 con $n=50$ y $n=100$ para el modelo de Cox con aproximadamente 50 %, 20 % y 30 % de datos completos, censurados por la derecha e izquierda, respectivamente, comparada con la gráfica (en rojo) de la densidad χ^2_1 .

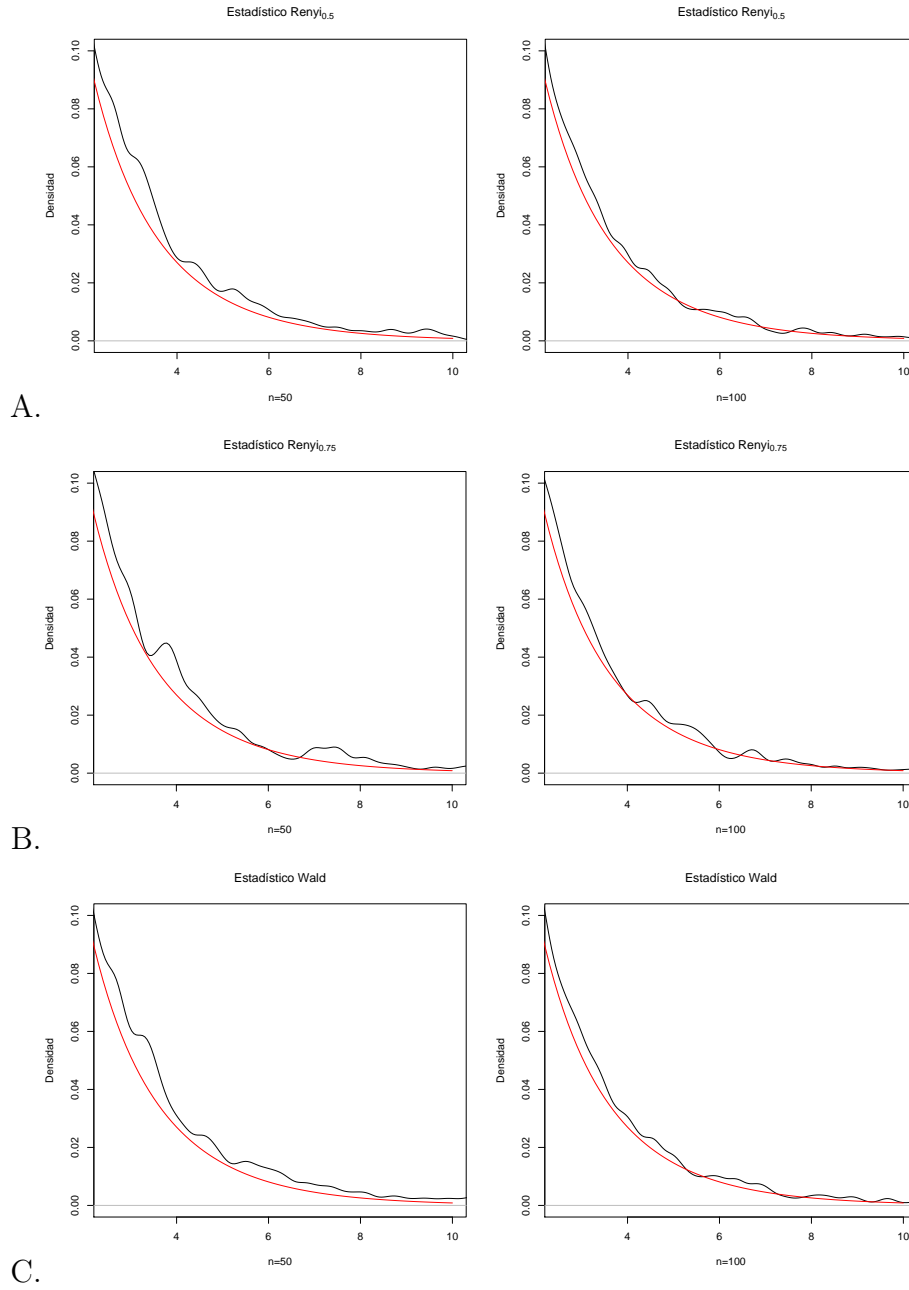


Figura 4.10: Gráfica de la densidad empírica de los estimadores de los estadísticos en las colas A. $Renyi_{0,50}$, B. $Renyi_{0,75}$ y C. $Wald$ bajo H_0 con $n=50$ y $n=100$ para el modelo de Cox con aproximadamente 50 %, 20 % y 30 % de datos completos, censurados por la derecha e izquierda, respectivamente, comparada con la gráfica (en rojo) de la densidad χ^2_1 .

a potencia se obtienen mejores resultados para el estadístico de $Renyi_{0,10}$ para n pequeño, lo cual es de esperar al tener un tamaño mayor. También se observa que para n grande la potencia asintótica se aproxima a la exacta. Todos estos valores se grafican en la Figura 4.11, en donde se puede ver que el estadístico LR , en general, tiene mejor comportamiento en cuanto Tamaño exacto - Tamaño nominal y Potencia exacta - Potencia asintótica, salvo en algunos casos con tamaño de muestra n moderado se tiene que el estadístico $Renyi_{0,10}$ es una mejor alternativa.

En las Figuras 4.12, 4.13 y 4.14 se muestran la estimaciones de la función de densidad de los estadísticos bajo H_0 , comparándolas con una distribución χ^2_1 . Es necesario decir, que con el estadístico de Wald se obtienen algunos valores demasiado grandes lo que hace que la gráfica de su densidad empírica tenga mucho sesgo, por lo anterior se ha omitido. En la Figura 4.15 se muestran las colas de ésta estimación para los tres mejores estadísticos que son $Kullback_1$, $Renyi_{0,1}$ y LR . De esta forma, el estadístico LR se prefiere para $\alpha = 0.05$ y todo n y para $\alpha = 0.10$ y valores pequeños de n , sin embargo para $\alpha = 0.10$ y valores moderados de n se prefiere el estadístico $Renyi_{0,1}$.

Inferencia para los Modelos

n	Estadístico	Tamaño exacto	Potencia exacta	Potencia asintótica	PE - TE
20	$Kullback_1$	0.0654	0.2550	0.5049	0.1896
	$Kullback_2$	0.1474	0.6124	0.7951	0.4650
	$Jeffreys$	0.1270	0.5220	0.7220	0.3950
	$Renyi_{0,10}$	0.0744	0.2984	0.5369	0.2240
	$Renyi_{0,25}$	0.0860	0.3546	0.5850	0.2686
	$Renyi_{0,50}$	0.1078	0.4464	0.6629	0.3386
	$Renyi_{0,75}$	0.1294	0.5328	0.7341	0.4034
	$Wald$	0.1674	0.6768	0.8390	0.5094
	LR	0.0550	0.2658	0.6684	0.2108
50	$Kullback_1$	0.0722	0.6280	0.8445	0.5558
	$Kullback_2$	0.0968	0.7754	0.9056	0.6786
	$Jeffreys$	0.0842	0.7220	0.8905	0.6378
	$Renyi_{0,10}$	0.0710	0.6420	0.8531	0.5710
	$Renyi_{0,25}$	0.0734	0.6636	0.8651	0.5902
	$Renyi_{0,50}$	0.0806	0.6982	0.8824	0.6176
	$Renyi_{0,75}$	0.0884	0.7372	0.8961	0.6488
	$Wald$	0.1116	0.8066	0.8982	0.6950
	LR	0.0648	0.5620	0.7393	0.4972
100	$Kullback_1$	0.0702	0.8288	0.9410	0.7586
	$Kullback_2$	0.0756	0.8986	0.9470	0.8230
	$Jeffreys$	0.0752	0.8700	0.9453	0.7948
	$Renyi_{0,10}$	0.0698	0.8362	0.9426	0.7664
	$Renyi_{0,25}$	0.0708	0.8446	0.9445	0.7738
	$Renyi_{0,50}$	0.0712	0.8620	0.9470	0.7908
	$Renyi_{0,75}$	0.0728	0.8784	0.9481	0.8056
	$Wald$	0.0860	0.9130	0.9309	0.8270
	LR	0.0564	0.8204	0.8541	0.7640
200	$Kullback_1$	0.0656	0.9586	0.9782	0.8930
	$Kullback_2$	0.0724	0.9876	0.9730	0.9152
	$Jeffreys$	0.0700	0.9756	0.9747	0.9056
	$Renyi_{0,10}$	0.0660	0.9610	0.9781	0.8950
	$Renyi_{0,25}$	0.0666	0.9654	0.9779	0.8988
	$Renyi_{0,50}$	0.0686	0.9720	0.9772	0.9034
	$Renyi_{0,75}$	0.0692	0.9806	0.9758	0.9114
	$Wald$	0.0752	0.9916	0.9575	0.9164
	LR	0.0522	0.9870	0.9459	0.9348

Tabla 4.5: Tamaño exacto, Potencia exacta, Potencia asintótica y Potencia exacta - Tamaño exacto para bases de datos ajustados al modelo generalizado con aproximadamente 25 % de censura por la derecha con $\alpha = 0.050$.

Inferencia para los Modelos

n	Estadístico	Tamaño exacto	Potencia exacta	Potencia asintótica	PE - TE
20	$Kullback_1$	0.1110	0.4624	0.6854	0.3514
	$Kullback_2$	0.1724	0.6902	0.8457	0.5178
	$Jeffreys$	0.1520	0.6262	0.8068	0.4742
	$Renyi_{0,10}$	0.1190	0.4876	0.7053	0.3686
	$Renyi_{0,25}$	0.1294	0.5250	0.7339	0.3956
	$Renyi_{0,50}$	0.1432	0.5808	0.7773	0.4376
	$Renyi_{0,75}$	0.1558	0.6366	0.8149	0.4808
	$Wald$	0.1866	0.7298	0.8599	0.5432
	LR	0.1224	0.4444	0.7527	0.3220
50	$Kullback_1$	0.1122	0.7362	0.8979	0.6240
	$Kullback_2$	0.1232	0.8420	0.9214	0.7188
	$Jeffreys$	0.1170	0.8054	0.9152	0.6884
	$Renyi_{0,10}$	0.1110	0.7478	0.9016	0.6368
	$Renyi_{0,25}$	0.1096	0.7658	0.9067	0.6562
	$Renyi_{0,50}$	0.1082	0.7908	0.9139	0.6826
	$Renyi_{0,75}$	0.1162	0.8194	0.9191	0.7032
	$Wald$	0.1334	0.8580	0.9070	0.7246
	LR	0.1228	0.7242	0.8261	0.6014
100	$Kullback_1$	0.1146	0.8932	0.9562	0.7786
	$Kullback_2$	0.1186	0.9442	0.9533	0.8256
	$Jeffreys$	0.1192	0.9256	0.9542	0.8064
	$Renyi_{0,10}$	0.1144	0.8978	0.9566	0.7834
	$Renyi_{0,25}$	0.1150	0.9078	0.9571	0.7928
	$Renyi_{0,50}$	0.1148	0.9210	0.9572	0.8062
	$Renyi_{0,75}$	0.1152	0.9342	0.9563	0.8190
	$Wald$	0.1230	0.9524	0.9352	0.8294
	LR	0.1084	0.9182	0.9088	0.8098
200	$Kullback_1$	0.1136	0.9918	0.9818	0.8782
	$Kullback_2$	0.1128	0.9988	0.9752	0.8860
	$Jeffreys$	0.1144	0.9966	0.9775	0.8822
	$Renyi_{0,10}$	0.1126	0.9934	0.9816	0.8808
	$Renyi_{0,25}$	0.1116	0.9950	0.9812	0.8834
	$Renyi_{0,50}$	0.1114	0.9960	0.9801	0.8846
	$Renyi_{0,75}$	0.1108	0.9974	0.9784	0.8866
	$Wald$	0.1154	0.9994	0.9594	0.8840
	LR	0.0948	0.9994	0.9692	0.9046

Tabla 4.6: Tamaño exacto, Potencia exacta, Potencia asintótica y Potencia exacta - Tamaño exacto para bases de datos ajustados al modelo generalizado con aproximadamente 25 % de censura por la derecha con $\alpha = 0.100$.

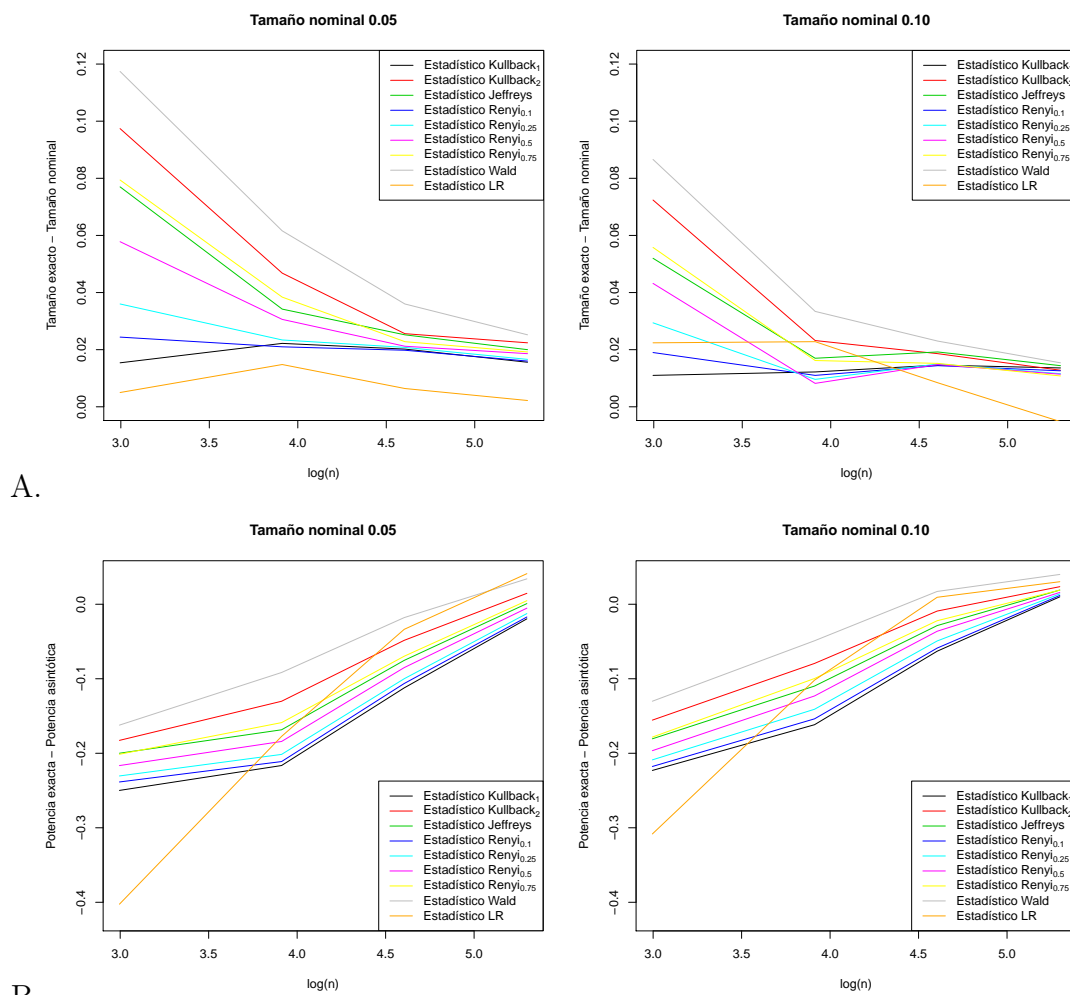


Figura 4.11: A. Gráfica tamaño exacto - tamaño nominal vs. $\log(n)$ y B. Gráfica potencia exacta - potencia asintótica vs. $\log(n)$ para el modelo generalizado de riesgos proporcionales con aproximadamente 25 % de censura por la derecha.

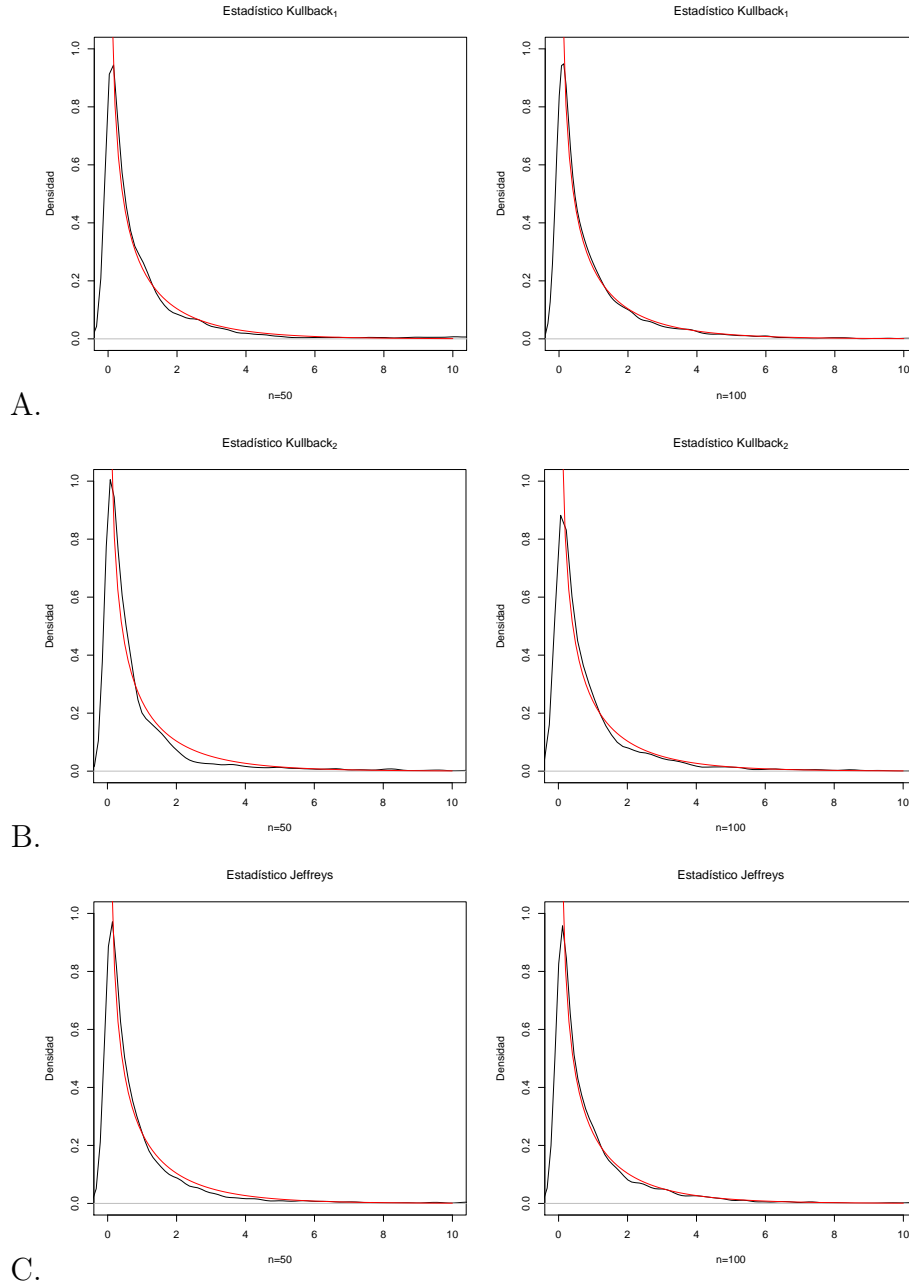


Figura 4.12: Gráfica de la densidad empírica de los estimadores de los estadísticos A. $Kullback_1$, B. $Kullback_2$ y C. $Jeffreys$ bajo H_0 con $n=50$ y $n=100$ para el modelo generalizado de riesgos proporcionales con aproximadamente 25 % de censura por la derecha comparada con la gráfica (en rojo) de la densidad χ^2_1 .

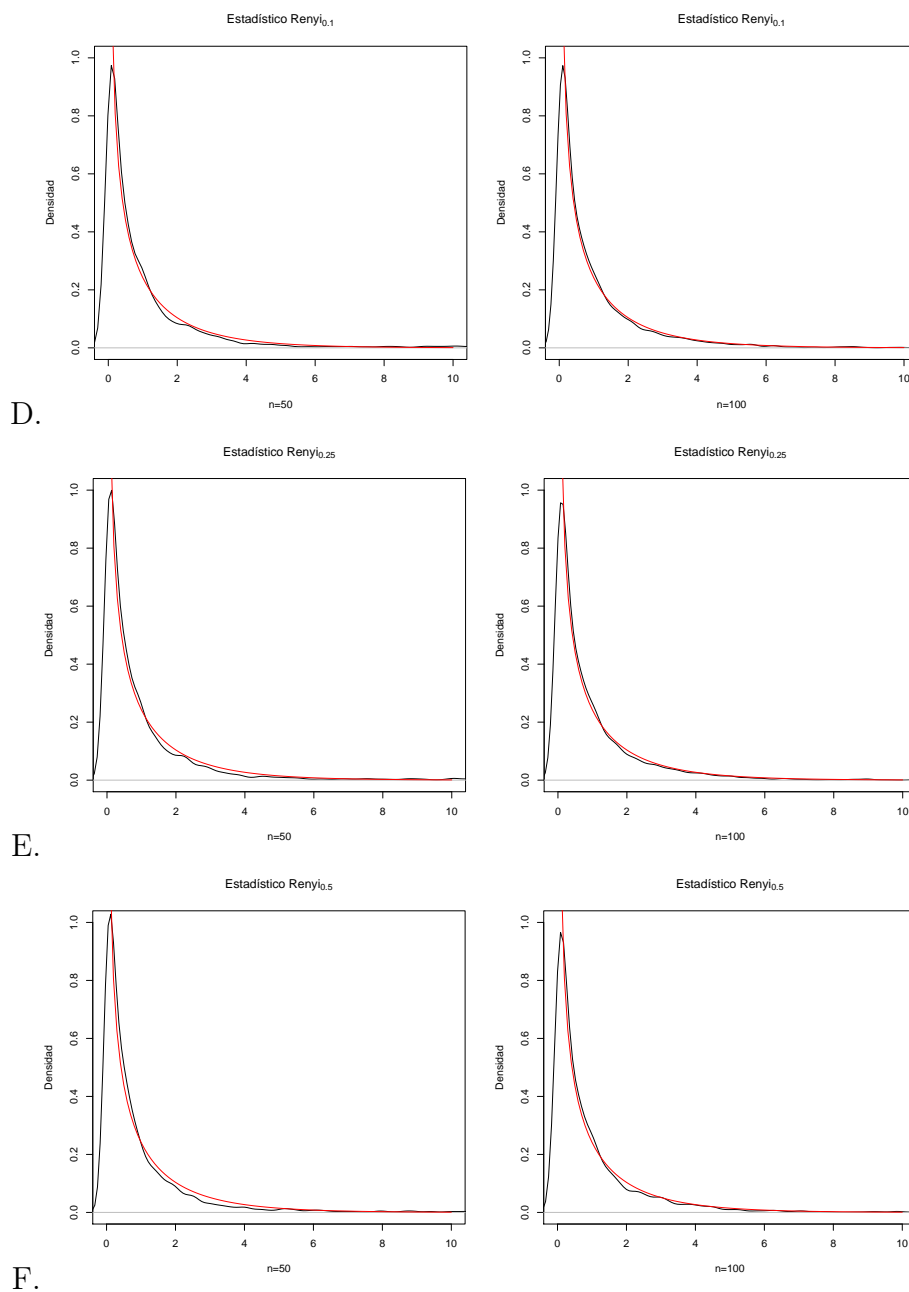


Figura 4.13: Gráfica de la densidad empírica de los estimadores de los estadísticos D. $Renyi_{0,10}$, E. $Renyi_{0,25}$ y F. $Renyi_{0,50}$ bajo H_0 con $n=50$ y $n=100$ para el modelo generalizado de riesgos proporcionales con aproximadamente 25 % de censura por la derecha comparada con la gráfica (en rojo) de la densidad χ^2_1 .

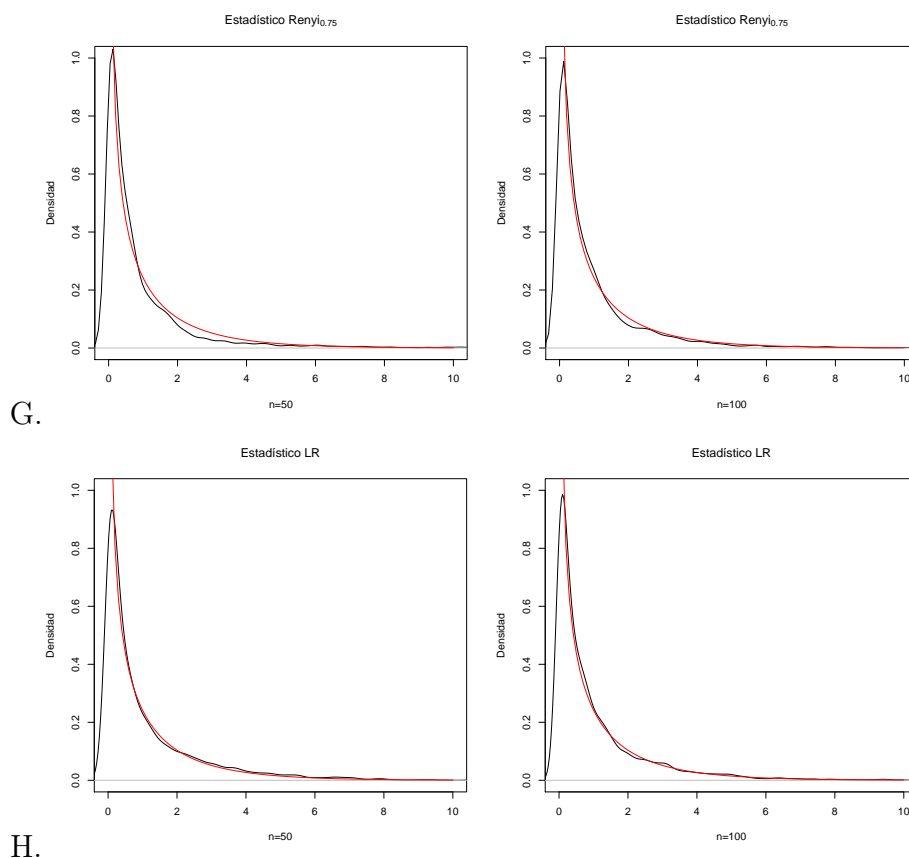


Figura 4.14: Gráfica de la densidad empírica de los estimadores de los estadísticos G. $Renyi_{0.75}$ y H. LR bajo H_0 con $n=50$ y $n=100$ para el modelo generalizado de riesgos proporcionales con aproximadamente 25 % de censura por la derecha comparada con la gráfica (en rojo) de la densidad χ^2_1 .

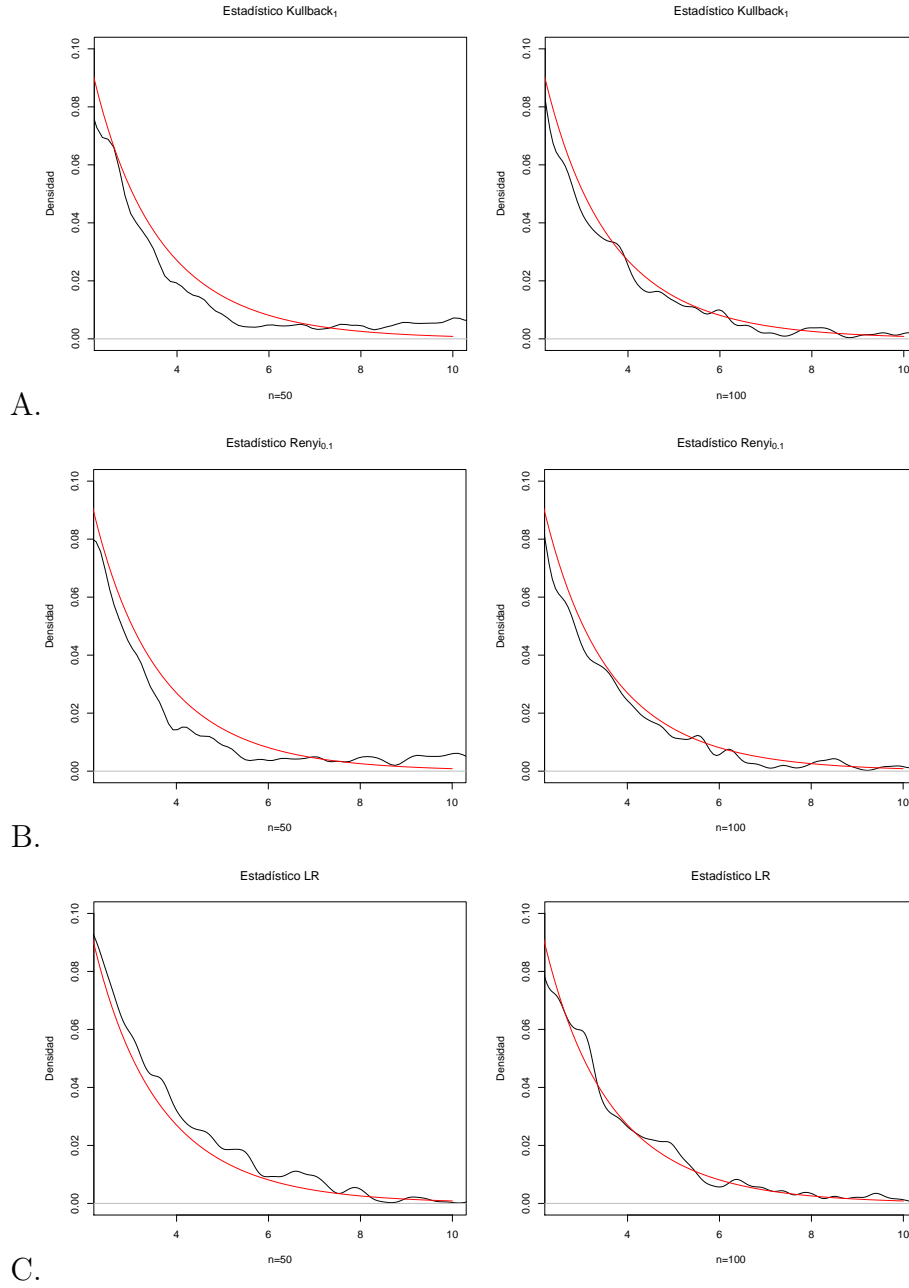


Figura 4.15: Gráfica de la densidad empírica de los estimadores de los estadísticos en las colas A. $Kullbak_1$, B. $Renyi_{0,10}$ y C. LR bajo H_0 con $n=50$ y $n=100$ para el modelo generalizado de riesgos proporcionales con aproximadamente 25 % de censura por la derecha comparada con la gráfica (en rojo) de la densidad χ^2_1 .

Capítulo 5

Conclusiones y Futuras Líneas de Investigación

5.1. Conclusiones

El modelo de Cox es, sin duda, uno de los modelos de riesgos proporcionales más conocido y utilizado. Sin embargo, otros modelos han surgido recientemente en la literatura como alternativa a éste. En esta tesis se estudia una nueva familia de modelos de riesgos proporcionales, cambiando la función nexo exponencial en el modelo de Cox por la función de distribución logística generalizada del Tipo-I. La nueva familia de modelos tiene un parámetro más que el modelo de Cox que aporta flexibilidad al modelo. Los diferentes conjuntos reales de datos estudiados ilustran dicha flexibilidad, ya que cuándo se interpretan los valores de los parámetros en base al riesgo relativo, se obtiene una interpretación más específica en comparación con el modelo de Cox. Además, los estimadores propuestos para los parámetros desconocidos del modelo no sólo tienen las propiedades asintóticas deseables, sino que el estudio de simulación llevado a cabo muestra su buen comportamiento para muestras finitas. Sin embargo, este comportamiento se ve influenciado no sólo por

Conclusiones y Futuras Líneas de Investigación

el tamaño de muestra sino también por la proporción de datos completos en el conjunto de datos. Por otro lado, en base del estimador de la función de riesgo base acumulativa para datos con censura por la derecha, se proponen tres estimadores de la función de supervivencia para la versión semiparamétrica del modelo. Estos estimadores se comparan mediante un estudio de simulación, del cual se concluye que uno de los estimadores basado en la definición de producto integral tiene mejor comportamiento que el resto cuando hay suficientes datos completos en la base de datos.

Por otra parte, no sólo para la censura más habitual que es la censura por la derecha se han estudiado nuevos modelos, sino que también para el modelo de Cox se ha considerado una censura más compleja pero que surge con frecuencia. En esta tesis se desarrolla toda la teoría necesaria para llevar a cabo la estimación del modelo de Cox para datos doblemente censurados y hacer inferencias sobre esos estimadores. Se demuestra la normalidad asintótica del estimador de los parámetros. Luego, se comprueba el buen comportamiento de los estimadores de los parámetros para muestras finitas en base a un estudio de simulación. Aunque este comportamiento se ve influenciado no sólo por el porcentaje de datos censurados, sino por la proporción de datos con censura por la izquierda o por la derecha, esto es, con menor porcentaje de datos con censura por la izquierda se tienen mejores resultados a pesar de que el porcentaje de datos completos sea el mismo. Además, se ilustra la teoría con un conjunto de datos reales con este tipo de censura. Se compara el riesgo relativo de dos individuos bajo la estimación de los parámetros del modelo de Cox ajustado a datos con censura por la derecha y ajustado a datos doblemente censurados y se observa que la diferencia de este riesgo es grande, lo cual indica que si sólo se usan los datos con censura por la derecha se pierde información valiosa para estimar el riesgo.

Finalmente, se proponen estadísticos basados en la divergencia de Kullback–

Leibler, de Jeffreys y de Rényi para hacer contrastes de hipótesis y verificar el efecto de las covariables, tanto para el modelo generalizado de riesgos proporcionales con datos censurados por la derecha como para el modelo de Cox con datos doblemente censurados. Se ha obtenido que la distribución asintótica de estos estadísticos, bajo la hipótesis nula, es una combinación lineal de distribuciones χ^2 con un grado de libertad, que permite construir los contrastes de hipótesis correspondientes. Se propone el uso de diferentes aproximaciones para calcular los puntos críticos de esa distribución, excepto en el caso unidimensional que la distribución asintótica se reduce a una χ^2 con un grado de libertad. También se obtiene la expresión de la potencia asintótica de estos estadísticos. Se ha llevado a cabo un estudio de simulación para comparar los estadísticos propuestos entre sí, además de compararlos con los estadísticos clásicos de Wald y cociente de verosimilitud. De este estudio se concluye que algunos miembros de la familia de estadísticos basada en la familia de divergencias de Rényi son una buena alternativa a los estadísticos clásicos para tamaños de muestra pequeños o, en algunos casos, moderados.

5.2. Futuras Líneas de Investigación

En cuanto a las líneas futuras de investigación en relación a esta tesis doctoral, destacan las siguientes:

1. En el modelo propuesto en el Capítulo 2 no se incluyen variables cuyos valores cambian con el tiempo. Sin embargo, el observar estas variables explicativas a intervalos regulares durante la duración del estudio proporciona una información muy valiosa. Así por ejemplo, en un estudio de comparación de 2 tratamientos para un determinado cáncer hay una serie de características, como puede ser el tamaño del tumor, que aparte de observarse al principio del estudio se deberían observar en diferentes tiempos del estudio, lo que lle-

varía a una mejor predicción de la esperanza de vida. Por ello, se pretende extender dicho modelo a un modelo de riesgo no proporcional. Lo cuál lleva a un nuevo modelo más flexible que contendría como caso particular el de MacKenzie (1997). A continuación, se desarrollaría la inferencia estadística asociada a este modelo, es decir, se obtendrían los estimadores de los parámetros desconocidos del modelo y sus propiedades asintóticas. Además se llevaría a cabo, la construcción de test de hipótesis relacionados. Finalmente, se ilustraría el modelo con conjuntos de datos reales y también se llevaría a cabo un estudio de simulación.

2. Se pretende buscar funciones nexo que engloben los nuevos modelos propuestos en el Capítulo 2 con algunos de los modelos ya conocidos de forma análoga a los trabajos de Younes y Lachin (1997) y Royston y Parmar (2002).
3. Por otro lado, el concepto de fragilidad (frailty) proporciona una forma adecuada de introducir efectos aleatorios, asociación y heterogeneidad no observada en modelos para datos de supervivencia. En su forma más simple, la fragilidad es un factor de proporcionalidad aleatorio no observable que modifica la función de riesgo de un individuo o de individuos relacionados. Este concepto apareció en el trabajo de Greenwood y Yule (1920), pero el término de fragilidad en sí fue introducido por Vaupel y otros (1979) en modelos de supervivencia univariantes y Clayton (1978) lo aplicó a datos multivariantes. No obstante, su popularización se logra desde finales de los años ochenta y principios de los noventa. Una breve introducción es presentada por Hougaard (1998) y Andersen y otros (1993). La extensión al caso multivariante se puede ver en Hougaard (2000). La idea es considerar modelos con fragilidad (frailty models) que extiendan el modelo introducido en el Capítulo 2 de tal forma que el riesgo de un individuo dependa además de una variable aleatoria no observable que interviene de forma multiplicativa en la función

Conclusiones y Futuras Líneas de Investigación

de riesgo base. Dentro de esta línea también se exploraría el uso en el modelo con fragilidad de otras funciones de riesgo base del tipo mixtura pensando en muestras de sujetos que consisten en 2 grupos y desarrollar la inferencia estadística correspondiente.

4. Una extensión natural del Capítulo 3 sería considerar un modelo generalizado de riesgos proporcionales, es decir cambiar la función nexa exponencial del modelo de Cox por otra función, como por ejemplo la función de distribución logística generalizada del Tipo-I para ajustar dicho modelo a datos doblemente censurados.
5. Con respecto al modelo de Cox ajustado a datos doblemente censurados, otra línea de investigación a seguir sería considerar variables explicativas que varían en el tiempo como las mencionadas en el punto 1 de este apartado. También sería interesante estudiar modelos con fragilidad como extensiones del modelo de Cox pero para datos doblemente censurados.
6. Aunque la teoría del Capítulo 4 se ha desarrollado para el caso de un vector de parámetros no unidimensional, el estudio de simulación se ha llevado a cabo sólo para el caso unidimensional. De igual forma, el contraste para el efecto de un subconjunto de las covariables ha sido propuesto teóricamente pero este caso no se ha incluido en el estudio de simulación. Sería interesante llevar a cabo un estudio de simulación considerando estos diferentes escenarios para comprobar si las conclusiones a las que se ha llegado en el caso unidimensional se mantienen.

Conclusiones y Futuras Líneas de Investigación

Summary

Survival analysis is a set of statistical techniques to analyze data corresponding to occurrence time of any event. In biomedicine the interest event usually is the death, that is, the analyzed time is the survival time.

The main objectives of survival analysis are to estimate the survival function, which is the probability to observe the interest event after any time period and to estimate the hazard function, which is the probability to observe the interest event at any time given it has survived to this time. The non-parametrics methods are useful for these estimations, but in several medical studies, additional information is available about the patient that affect survival time. In this case, the main aim of survival analysis is to propose a model for the hazard function in presence of explanatory variables.

The aim of this thesis is twofold. Firstly, to develop new models of survival analysis more flexible than existing for right censored data, that is the most often type of censoring present in the real data and therefore one of the most studied in the literature. Secondly, to study known survival models for complex mechanisms of censoring. In both cases the necessary theory to perform statistical inference will be developed, and also, some simulation studies will be made.

Regarding the first objective, a new family of proportional hazards models is proposed changing the link exponential function in the Cox's model by the generalized logistic distribution function Type-I. This generalized proportional hazards

model is studied for right censored survival data. It should be emphasized that the new family of models has one parameter more than the traditional Cox's model, this extra parameter gives flexibility to the model. In fact, more precise conclusions are obtained interpreting the relative risk associated with this new model compared with the Cox's model. The parametric and semiparametric versions of this new family of models are studied. The consistency and the asymptotic normality of the parameter estimators of the model are established and different estimators of the survival function are proposed. In particular, for the semi-parametric model, three estimators of the survival function are proposed based on the estimator of the cumulative baseline hazard function. The estimator of the cumulative baseline hazard function is the equivalent to the Breslow's estimator in the Cox's model. A simulation study, for small and moderate samples, is carried out on the estimates of the covariates effects as well as the estimators of the survival function. Furthermore, this model is illustrated with two datasets of survival data of cancer, one of them about laryngeal cancer and other one about lung cancer. Also, test statistics to test hypotheses about the effects of covariates are proposed based on the Kullback–Leibler, Rényi and Jeffreys divergences. The asymptotic distribution of these test statistics under the null hypothesis is obtained to build the critical region of the corresponding hypothesis test, and their asymptotic power is also obtained. The proposed test statistics are compared through a simulation study with the Wald and likelihood ratio classical test statistics.

Regarding the second objective, a regression model for doubly censored data is proposed. Doubly censored data have been studied by several authors who have proposed estimators of the survival function for them, however models with explanatory variables have not been proposed. In this work, the doubly censored data are modeled through latent variables and covariates are considered. An estimator of the cumulative baseline hazard function is defined using latent variables. Having done

this, the parameters of Cox's model are estimated maximizing the profile likelihood function. Moreover, the consistency and asymptotic normality of these estimators are established. Subsequently, through simulation studies, the behavior of the estimators for small or moderate samples is studied. After that, the Cox's model is fitted to doubly censored real data, these data are from a study about the AIDS presence in HIV-infected women. Finally, test statistics to make hypotheses test for the effect of covariates are proposed, they are based on the Kullback–Leibler, Rényi and Jeffreys divergences. The asymptotic distribution of the test statistics is established in order to build the critical region of the hypothesis test. Also, the asymptotic power of these statistics is obtained. Finally, a simulation study is performed to compare the proposed statistics and the Wald and likelihood ratio classical test statistics. Although these test statistics have asymptotically the same distribution but their behavior, for small or moderate samples, are not necessary the same. The test statistics are compared in terms of their power and their size.

Undoubtedly, the Cox's model is one of the most known and useful proportional hazards models in survival analysis. However, recently other models have emerged in the literature as alternative to it. In this thesis, a generalized proportional hazards model is studied, which allows, from the estimated relative risk, a more specific interpretation for parameter values is obtained than that obtained with the Cox's model which is illustrated with the two different real datasets. The simulation study carried out shows the good behavior of the proposed estimators of the unknown parameters of the generalized proportional hazards model for finite samples. However, this behavior is influenced by the sample size and by the rate of complete data in the dataset. The proposed estimators of the survival function for the semiparametric generalized proportional hazards model based on the estimate of the cumulative baseline hazard function for right censored data are compared through a simulation study. We conclude that one of the estimators based on the

integrated product definition has better performance than the others when there are enough complete data in the dataset.

On the other hand, not only all the asymptotic theory is developed to estimate the unknown parameters of the Cox's model for doubly censored data but also the good performance of the parameters estimators in finite samples is verified through a simulation study. Although this behavior is influenced by the percentage of censored data, and by the rate of left or right censored data. In fact, the outperformance is better with lower percentage of left censored data even though the percentage of complete data is the same. Moreover, a real dataset with this type of censoring illustrates the theoretical results.

Finally, test statistics for hypothesis testing to verify the effect of covariates based on the Kullback–Leibler, Jeffreys and Rényi divergences are proposed for both models, the generalized proportional hazards model with right censored data and the Cox model with doubly censored data. The asymptotic distribution of these test statistics under the null hypothesis is a linear combination of distributions χ^2 with one degree of freedom, which allows to build the corresponding hypothesis test. Different approaches to calculate the critical points of this distribution are proposed, except from the unidimensional case, since the asymptotic distribution is a χ^2 with one degree of freedom in this case. Moreover, the expression of the asymptotic power of these test statistics is obtained. It has been carried out a simulation study to compare the proposed test statistics among them and with the Wald and likelihood ratio classical test statistics. This study concludes that some members of the family of statistics based on the Rényi family divergence are a good alternative to the classical test statistics for small sample sizes or, in some cases, moderate sample sizes.

Apéndice I: Derivadas

A. Modelo Paramétrico Generalizado para Datos con Censura por la Derecha (Sección 2.5)

Como se menciona en la Sección 2.5, para obtener los parámetros del modelo (2.5.7) se usan métodos de optimización numérica para maximizar la función (2.5.9) y dichos métodos suelen usar el gradiente de esta función. A continuación se escriben las expresiones de las derivadas parciales de $\ell_n(\mu, \gamma, \boldsymbol{\beta}, a)$ con respecto a cada parámetro del modelo.

$$\frac{\partial \ell_n(\mu, \gamma, \boldsymbol{\beta}, a)}{\partial \mu} = \sum_{i=1}^n \left[\frac{\delta_i}{\mu} - K(\boldsymbol{\beta}, a, \mathbf{z}_i) x_i^\gamma \right],$$

$$\frac{\partial \ell_n(\mu, \gamma, \boldsymbol{\beta}, a)}{\partial \gamma} = \sum_{i=1}^n \left[\delta_i \left(\frac{1}{\gamma} + \log(x_i) \right) - \mu K(\boldsymbol{\beta}, a, \mathbf{z}_i) x_i^\gamma \log(x_i) \right],$$

$$\frac{\partial \ell_n(\mu, \gamma, \boldsymbol{\beta}, a)}{\partial \beta_j} = \sum_{i=1}^n \left[\frac{a z_{ij} (\delta_i - \mu K(\boldsymbol{\beta}, a, \mathbf{z}_i) x_i^\gamma)}{1 + \exp(\boldsymbol{\beta}' \mathbf{z}_i)} \right],$$

Apéndice I: Derivadas

para $j = 1, \dots, d$, y

$$\frac{\partial \ell_n(\mu, \gamma, \boldsymbol{\beta}, a)}{\partial a} = \sum_{i=1}^n [\log(1 + \exp(-\boldsymbol{\beta}' \mathbf{z}_i)) (\mu K(\boldsymbol{\beta}, a, \mathbf{z}_i) x_i^\gamma - \delta_i)].$$

Por otro lado, para obtener la respectiva matriz de información \mathbf{I} del modelo, es necesario tener la matriz Hessiana de la función (2.5.9), lo cual implica obtener las segundas derivadas de la función (2.5.9) con respecto a cada parámetro, las cuales vienen dadas por

$$\frac{\partial^2 \ell_n(\mu, \gamma, \boldsymbol{\beta}, a)}{\partial \mu^2} = -\frac{1}{\mu^2} \left[\sum_{i=1}^n \delta_i \right],$$

$$\frac{\partial^2 \ell_n(\mu, \gamma, \boldsymbol{\beta}, a)}{\partial \mu \partial \gamma} = -\sum_{i=1}^n K(\boldsymbol{\beta}, a, \mathbf{z}_i) x_i^\gamma \log(x_i),$$

$$\frac{\partial^2 \ell_n(\mu, \gamma, \boldsymbol{\beta}, a)}{\partial \mu \partial \beta_j} = -a \sum_{i=1}^n \frac{K(\boldsymbol{\beta}, a, \mathbf{z}_i) x_i^\gamma z_{ij}}{1 + \exp(\boldsymbol{\beta}' \mathbf{z}_i)}, \quad \text{con } j = 1, \dots, d,$$

$$\frac{\partial^2 \ell_n(\mu, \gamma, \boldsymbol{\beta}, a)}{\partial \mu \partial a} = \sum_{i=1}^n K(\boldsymbol{\beta}, a, \mathbf{z}_i) x_i^\gamma \log(1 + \exp(-\boldsymbol{\beta}' \mathbf{z}_i)),$$

$$\frac{\partial^2 \ell_n(\mu, \gamma, \boldsymbol{\beta}, a)}{\partial \gamma^2} = \sum_{i=1}^n \left[-\frac{\delta_i}{\gamma^2} - \mu K(\boldsymbol{\beta}, a, \mathbf{z}_i) x_i^\gamma (\log(x_i))^2 \right],$$

$$\frac{\partial^2 \ell_n(\mu, \gamma, \boldsymbol{\beta}, a)}{\partial \gamma \partial \beta_j} = -a \mu \sum_{i=1}^n \frac{K(\boldsymbol{\beta}, a, \mathbf{z}_i) x_i^\gamma \log(x_i) z_{ij}}{1 + \exp(\boldsymbol{\beta}' \mathbf{z}_i)}, \quad \text{con } j = 1, \dots, d,$$

$$\frac{\partial^2 \ell_n(\mu, \gamma, \boldsymbol{\beta}, a)}{\partial \gamma \partial a} = \mu \sum_{i=1}^n K(\boldsymbol{\beta}, a, \mathbf{z}_i) x_i^\gamma \log(x_i) \log(1 + \exp(-\boldsymbol{\beta}' \mathbf{z}_i)),$$

para $j = 1, \dots, d$ y $l = j, \dots, d$,

$$\begin{aligned} \frac{\partial^2 \ell_n(\mu, \gamma, \boldsymbol{\beta}, a)}{\partial \beta_j \partial \beta_l} = & -a \sum_{i=1}^n \left[\frac{\delta_i \exp(\boldsymbol{\beta}' \mathbf{z}_i) z_{ij} z_{il}}{(1 + \exp(\boldsymbol{\beta}' \mathbf{z}_i))^2} + \right. \\ & \left. \frac{\mu K(\boldsymbol{\beta}, a, \mathbf{z}_i) x_i^\gamma z_{ij} z_{il} (a - \exp(\boldsymbol{\beta}' \mathbf{z}_i))}{(1 + \exp(\boldsymbol{\beta}' \mathbf{z}_i))^2} \right], \end{aligned}$$

$$\frac{\partial^2 \ell_n(\mu, \gamma, \boldsymbol{\beta}, a)}{\partial \beta_j \partial a} = \sum_{i=1}^n \left[\frac{\delta_i z_{ij} - \mu K(\boldsymbol{\beta}, a, \mathbf{z}_i) t_i^\gamma z_{ij} \{1 - a \log(1 + \exp(-\boldsymbol{\beta}' \mathbf{z}_i))\}}{1 + \exp(\boldsymbol{\beta}' \mathbf{z}_i)} \right],$$

$$\frac{\partial^2 \ell_n(\mu, \gamma, \boldsymbol{\beta}, a)}{\partial a^2} = -\mu \sum_{i=1}^n K(\boldsymbol{\beta}, a, \mathbf{z}_i) x_i^\gamma \{\log(1 + \exp(-\boldsymbol{\beta}' \mathbf{z}_i))\}^2.$$

B. Modelo Semiparamétrico Generalizado para Datos Censurados por la Derecha (Sección 2.6)

Al igual que en el modelo paramétrico, para obtener los parámetros del modelo semiparamétrico se debe maximizar numéricamente la respectiva función de log-verosimilitud (2.6.11), para ello es necesario obtener su función gradiente, por tanto, se escriben las derivadas parciales de $\ell(\boldsymbol{\beta}, a)$ como

$$\frac{\partial \ell_n(\boldsymbol{\beta}, a)}{\partial \beta_j} = a \sum_{i=1}^n \delta_i \left[\frac{z_{ij}}{1 + \exp(\boldsymbol{\beta}' \mathbf{z}_i)} - \frac{1}{\sum_{l \in R(x_i)} K(\boldsymbol{\beta}, a, \mathbf{z}_l)} \sum_{l \in R(x_i)} \frac{z_{lj} K(\boldsymbol{\beta}, a, \mathbf{z}_l)}{1 + \exp(\boldsymbol{\beta}' \mathbf{z}_l)} \right]$$

Apéndice I: Derivadas

para $j = 1, \dots, d$, y

$$\begin{aligned} \frac{\partial \ell_n(\boldsymbol{\beta}, a)}{\partial a} = & \sum_{i=1}^n \delta_i \left[-\log(1 + \exp(-\boldsymbol{\beta}' \mathbf{z}_i)) \right. \\ & \left. + \frac{1}{\sum_{l \in R(x_i)} K(\boldsymbol{\beta}, a, \mathbf{z}_l)} \sum_{l \in R(x_i)} K(\boldsymbol{\beta}, a, \mathbf{z}_l) \log(1 + \exp(-\boldsymbol{\beta}' \mathbf{z}_l)) \right]. \end{aligned}$$

Además, para obtener la matriz de información del modelo, se requieren las segundas derivadas parciales de la función de log-verosimilitud (2.6.11) para tener su matriz Hessiana, las cuales se escriben para $j = 1, \dots, d$ y $m = j, \dots, d$ como sigue:

$$\begin{aligned} \frac{\partial^2 \ell_n(\boldsymbol{\beta}, a)}{\partial \beta_j \partial \beta_m} &= a \sum_{i=1}^n \delta_i \left[\frac{-\exp(\boldsymbol{\beta}' \mathbf{z}_i) z_{ij} z_{im}}{(1 + \exp(\boldsymbol{\beta}' \mathbf{z}_i))^2} - \frac{av_m(x_i) v_j(x_i) - w(x_i) y_{jm}(x_i)}{(w(x_i))^2} \right], \\ \frac{\partial^2 \ell_n(\boldsymbol{\beta}, a)}{\partial \beta_j \partial a} &= \sum_{i=1}^n \delta_i \left[\frac{z_{ij}}{1 + \exp(\boldsymbol{\beta}' \mathbf{z}_i)} - \frac{v_j(x_i)}{w(x_i)} - a \frac{v_j(x_i) u(x_i)}{(w(x_i))^2} + a \frac{z_j(x_i)}{w(x_i)} \right], \\ \frac{\partial^2 \ell_n(\boldsymbol{\beta}, a)}{\partial a^2} &= \sum_{i=1}^n \delta_i \left[\frac{-(u(x_i))^2}{(w(x_i))^2} + \frac{\sum_{l \in R(x_i)} K(\boldsymbol{\beta}, a, \mathbf{z}_l) \{\log(1 + \exp(-\boldsymbol{\beta}' \mathbf{z}_l))\}^2}{w(x_i)} \right], \end{aligned}$$

donde

$$\begin{aligned}
 w(x) &= \sum_{l \in R(x)} K(\boldsymbol{\beta}, a, \mathbf{z}_l), \\
 v_j(x) &= \sum_{l \in R(x)} \frac{K(\boldsymbol{\beta}, a, \mathbf{z}_l) z_{lj}}{1 + \exp(\boldsymbol{\beta}' \mathbf{z}_l)}, \\
 y_{jm}(x) &= \sum_{l \in R(x)} \frac{K(\boldsymbol{\beta}, a, \mathbf{z}_l) z_{lj} z_{lm} (a - \exp(\boldsymbol{\beta}' \mathbf{z}_l))}{(1 + \exp(\boldsymbol{\beta}' \mathbf{z}_l))^2}, \\
 u(x) &= \sum_{l \in R(x)} K(\boldsymbol{\beta}, a, \mathbf{z}_l) \log(1 + \exp(-\boldsymbol{\beta}' \mathbf{z}_l)), \\
 z_j(x) &= \sum_{l \in R(x)} \frac{K(\boldsymbol{\beta}, a, \mathbf{z}_l) z_{lj} \{\log(1 + \exp(-\boldsymbol{\beta}' \mathbf{z}_l))\}}{1 + \exp(\boldsymbol{\beta}' \mathbf{z}_l)}.
 \end{aligned}$$

C. Modelo de Cox para Datos Doblemente Censurados (Capítulo 3)

Para obtener los parámetros del modelo de Cox (3.3.3) ajustado a un conjunto de datos doblemente censurados es necesario maximizar la función de log-verosimilitud (3.3.15), para ello se usan métodos de optimización numérica del tipo Newton-Raphson que requieren la función gradiente de (3.3.15). A continuación, se escribe la expresión de la función derivada parcial de (3.3.15) con respecto a $\boldsymbol{\beta}_\ell$ con $\ell = 1, \dots, d$.

$$\begin{aligned}
 \frac{\partial \ell_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_\ell} &= \sum_{i=1}^n \Delta N_{0i} \left(\mathbf{z}_i - \frac{E^{(1)}(x_i | \boldsymbol{\beta}, \hat{p})}{E^{(0)}(x_i | \boldsymbol{\beta}, \hat{p})} - \frac{E^{(1)}(x_i | \boldsymbol{\beta}, 1)}{E^{(0)}(x_i | \boldsymbol{\beta}, 1)} + \frac{E^{(0)}(x_i | \boldsymbol{\beta}, 1) E^{(1)}(x_i | \boldsymbol{\beta}, \hat{p})}{[E^{(0)}(x_i | \boldsymbol{\beta}, \hat{p})]^2} \right) \\
 &+ \sum_{i=1}^n \Delta N_{2i} \left(\frac{\exp(-v(x_i | \boldsymbol{\beta}, \hat{p}))}{1 - \exp(-v(x_i | \boldsymbol{\beta}, \hat{p}))} w(x_i | \boldsymbol{\beta}, \hat{p}) \right)
 \end{aligned}$$

donde

$$v(x_i|\boldsymbol{\beta}, \hat{p}) = \sum_{j=1}^i \frac{\exp(\boldsymbol{\beta}' \mathbf{z}_j)}{E^{(0)}(x_j|\boldsymbol{\beta}, \hat{p})} \Delta N_{0j}$$

y

$$\begin{aligned} w(x_i|\boldsymbol{\beta}, \hat{p}) &= \frac{\partial}{\partial \boldsymbol{\beta}_\ell} v(x_i|\boldsymbol{\beta}, \hat{p}) \\ &= \sum_{j=1}^i \Delta N_{0j} \exp(\boldsymbol{\beta}' \mathbf{z}_j) \left(\frac{\mathbf{z}_j}{E^{(0)}(x_j|\boldsymbol{\beta}, \hat{p})} - \frac{E^{(1)}(x_j|\boldsymbol{\beta}, \hat{p})}{[E^{(0)}(x_j|\boldsymbol{\beta}, \hat{p})]^2} \right). \end{aligned}$$

Apéndice II: Integración por Partes

En este apéndice se ilustra la integración por partes utilizada en varias ocasiones en la Tesis. Se define

$$u(x) = \left[\int_0^x h(s) d\Lambda_0(s) \right]$$

y

$$v(x) = \left[\int_0^x g(s) d\Lambda_0(s) \right],$$

entonces

$$\begin{aligned} & \int_0^x h(s) d\Lambda_0(s) \left[\int_0^x g(s) d\Lambda_0(s) \right] \\ &= u(x)v(x) \\ &= \int_0^x u(s) dv(s) + \int_0^x v(s) du(s) \\ &= \int_0^x \left[\int_0^s h(u) d\Lambda_0(u) \right] g(s) d\Lambda_0(s) + \int_0^x \left[\int_0^s g(u) d\Lambda_0(u) \right] h(s) d\Lambda_0(s) \\ &= \int_0^x \left[g(s) \int_0^s h(u) d\Lambda_0(u) + h(s) \int_0^s g(u) d\Lambda_0(u) \right] d\Lambda_0(s) \\ &= \int_0^x a(s) d\Lambda_0(s) \end{aligned}$$

con

$$a(s) = g(s) \int_0^s h(u) d\Lambda_0(u) + h(s) \int_0^s g(u) d\Lambda_0(u).$$

Apéndice II: Integración por Partes

Apéndice III: Bases de Datos

A. Cáncer de Laringe

	Etapa de la enfermedad	Tiempo de censura o supervivencia	Edad del paciente	Año de diagnóstico	Delta
1	1	0.6	77	76	1
2	1	1.3	53	71	1
3	1	2.4	45	71	1
4	1	2.5	57	78	0
5	1	3.2	58	74	1
6	1	3.2	51	77	0
7	1	3.3	76	74	1
8	1	3.3	63	77	0
9	1	3.5	43	71	1
10	1	3.5	60	73	1
11	1	4.0	52	71	1
12	1	4.0	63	76	1
13	1	4.3	86	74	1
14	1	4.5	48	76	0
15	1	4.5	68	76	0

Apéndice III: Bases de Datos

	Etapas de la enfermedad	Tiempo de censura o supervivencia	Edad del paciente	Año de diagnóstico	Delta
16	1	5.3	81	72	1
17	1	5.5	70	75	0
18	1	5.9	58	75	0
19	1	5.9	47	75	0
20	1	6.0	75	73	1
21	1	6.1	77	75	0
22	1	6.2	64	75	0
23	1	6.4	77	72	1
24	1	6.5	67	70	1
25	1	6.5	79	74	0
26	1	6.7	61	74	0
27	1	7.0	66	74	0
28	1	7.4	68	71	1
29	1	7.4	73	73	0
30	1	8.1	56	73	0
31	1	8.1	73	73	0
32	1	9.6	58	71	0
33	1	10.7	68	70	0
34	2	0.2	86	74	1
35	2	1.8	64	77	1
36	2	2.0	63	75	1
37	2	2.2	71	78	0
38	2	2.6	67	78	0
39	2	3.3	51	77	0
40	2	3.6	70	77	1

Apéndice III: Bases de Datos

	Etapas de la enfermedad	Tiempo de censura o supervivencia	Edad del paciente	Año de diagnóstico	Delta
41	2	3.6	72	77	0
42	2	4.0	81	71	1
43	2	4.3	47	76	0
44	2	4.3	64	76	0
45	2	5.0	66	76	0
46	2	6.2	74	72	1
47	2	7.0	62	73	1
48	2	7.5	50	73	0
49	2	7.6	53	73	0
50	2	9.3	61	71	0
51	3	0.3	49	72	1
52	3	0.3	71	76	1
53	3	0.5	57	74	1
54	3	0.7	79	77	1
55	3	0.8	82	74	1
56	3	1.0	49	76	1
57	3	1.3	60	76	1
58	3	1.6	64	72	1
59	3	1.8	74	71	1
60	3	1.9	72	74	1
61	3	1.9	53	74	1
62	3	3.2	54	75	1
63	3	3.5	81	74	1
64	3	3.7	52	77	0
65	3	4.5	66	76	0

Apéndice III: Bases de Datos

	Etapa de la enfermedad	Tiempo de censura o supervivencia	Edad del paciente	Año de diagnóstico	Delta
66	3	4.8	54	76	0
67	3	4.8	63	76	0
68	3	5.0	59	73	1
69	3	5.0	49	76	0
70	3	5.1	69	76	0
71	3	6.3	70	72	1
72	3	6.4	65	72	1
73	3	6.5	65	74	0
74	3	7.8	68	72	1
75	3	8.0	78	73	0
76	3	9.3	69	71	0
77	3	10.1	51	71	0
78	4	0.1	65	72	1
79	4	0.3	71	76	1
80	4	0.4	76	77	1
81	4	0.8	65	76	1
82	4	0.8	78	77	1
83	4	1.0	41	77	1
84	4	1.5	68	73	1
85	4	2.0	69	76	1
86	4	2.3	62	71	1
87	4	2.9	74	78	0
88	4	3.6	71	75	1
89	4	3.8	84	74	1
90	4	4.3	48	76	0

Bibliografía

- Andersen, P.K., Borgan, Φ ., Gill, R.D. y Keiding, N. (1993). *Statistical models based on counting processes*. Springer, New York.
- Andersen, P.K., Bentzon, M.W. y Klein, J.P. (1996). Estimating the survival function in the proportional hazards regression model: A study of the small sample size properties. *Scandinavian Journal of Statistics*, 23:1–12.
- Andersen, P.K. y Gill, R.D. (1982). Cox's regression model for counting processes: A large sample study. *Annals on Statistics*, 10(4):1100–1120.
- Balakrishnan, N. (Ed.) (1992). *Handbook of the logistic distribution*. Marcel Dekker, New York.
- Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35:99–109.
- Bender, R., Augustin, T. y Blettner, M. (2005). Generating Survival Times to Simulate Cox Proportional Hazards Model. *Statistics in Medicine*, 24(11):1713–1723.
- Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30:89–99.

Bibliografía

- Chang, M.N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Annals of Statistics*, 18:391–404.
- Chang, M.N. y Yang, G.L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *Annals of Statistics*, 15:1536–1547.
- Chen, H.S., Lai, K. y Ying, Z. (2004). Goodness-of-fit tests and minimum power divergence estimators for survival data. *Statistica Sinica*, 14:231–248.
- Chen, K. y Zhou, M. (2003). Nonparametric hypothesis testing and confidence intervals with doubly censored data. *Lifetime Data Analysis*, 9:71–91.
- Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65:141–151.
- Cox, D.R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, 34(2):187–220.
- Cox, D.R. (1975). Partial likelihood. *Biometrika*, 62:269–276.
- Cressie, N. y Read, T.R. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B*, 46(3):440–464.
- De Gruttola, V. y Lagakos, S.W. (1989). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics*, 45:1–11.
- Devarajan, K. y Ebrahimi, N. (2009). Testing for covariate effect in the Cox proportional hazards regression model. *Communication in Statistics Theory and Methods*, 38:2333–2347.
- Dik, J.J. y Gunst, M.C.M. (1985). The distribution of general quadratic forms in normal variables. *Statistica Neerlandica*, 39:14–26.

Bibliografía

- Eckler, A.R. (1969). A survey of coverage problems associated with point and area targets. *Technometrics*, 11:561–589.
- Etezadi-Amoli, J. y Ciampi, A. (1987). Extended hazard regression for censored survival data with covariates: A spline approximation for the background hazard function. *Biometrics*, 43:181–192.
- Gill, R.D. (1984). Understanding Cox’s regression model. A martingale approach. *Journal of the American Statistical Association*, 79:441–447.
- Gómez, G. y Lagakos, S.W. (1994). Estimation of the infection time and latency distribution of AIDS with doubly censored data. *Biometrics*, 50(1):204–212.
- Greenwood, M. y Yule, G.U. (1920). An inquiry into the nature of frequency distributions of multiple happenings, with particular reference to the occurrence of multiple attacks of disease or repeated accidents. *Journal of the Royal Statistical Society A*, 83:255–279.
- Groeneboom, P. y Wellner, J.A. (1992). *Information bounds and nonparametric maximum likelihood estimation*. Birkhäuser, Basel.
- Gu, M.G. y Zhang, C.H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Annals of Statistics*, 21:611–624.
- Gupta, S.S. (1963). Bibliography on the multivariate normal integrals and related topics. *Annals of Mathematical Statistics*, 34:829–838.
- Hald, A. (1949). Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. *Scandinavian Actuarial Journal*, 32:119–134.

Bibliografia

- Hirose, Y. (2010). Efficiency of profile/partial likelihood in the Cox model. *Research Report Series, School of Mathematics, Statistics and Operations Research*, Victoria University of Wellington, ISSN:1174–2011.
- Hougaard, P. (1998). *Frailty*, En: Armitage, P. y Colton, T. (Eds.), *Encyclopedia of Biostatistics*, John Wiley and Sons, New York, 1572–1577.
- Hougaard, P. (2000). *Analysis of multivariate survival data*. Springer, New York.
- Huang, J. (1999). Asymptotic properties of nonparametric estimation based on partly interval-censored data. *Statistica Sinica*, 9:501–519.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society, Series A*, 186:453–461.
- Johansen, S. (1983). An extension of Cox's regression model. *International Statistical Review*, 51(2):165–174.
- Johnson, N.L. y Kotz, S. (1968). Tables of distributions of positive definite quadratic forms in central normal variables. *Sankhya, Series B*, 30:303–314.
- Johnson, M.E., Tolley, H.D., Bryson, M.C. y Goldman, A.S. (1982). Covariate analysis of survival data: a small-sample study of Cox's model. *Biometrics*, 38:685–698.
- Julià, O. y Gómez G. (2011). Simultaneous marginal survival estimators when doubly censored data is present. *Life Data Analysis*, 17:347–372.
- Kalbfleish, J.D. y Prentice, R.L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika*, 60:267–278.
- Kalbfleish, J.D. y Prentice, R.L. (1980). *The statistical analysis of failure time data*. John Wiley and Sons, New York.

Bibliografía

- Kardaun, O. (1983). Statistical analysis of male larynx-cancer patients: A case study. *Statistical Nederlandica*, 37:103–126.
- Kim, M.Y., De Gruttola, V.G. y Lagakos, S.W. (1993). Analyzing doubly censored data with covariates, with application to AIDS. *Biometrics*, 49(1):13–22.
- Kim, J.S. (2003). Maximum likelihood estimation for the proportional hazards models with partly interval-censored data. *Journal of the Royal Statistical Society, Series B*, 65:489–502.
- Klein, J.P. y Moeschberger, M.L. (2003). *Survival analysis: Techniques for censored and truncated data*. Springer-Verlag, New York.
- Kosorok, M.R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer.
- Kullback, S. y Leibler, R.A. (1951). On Information and Sufficiency. *Annals of Mathematical Statistics*, 22 (1):79–86.
- Kullback, S. (1959). *Information theory and statistics*. John Wiley and Sons, New York.
- Lau, B.M., Cole, S.R. y Gange, S.J. (2009). Competing risk regression models for epidemiologic data. *American Journal of Epidemiology*, 170(2):244–56.
- MacKenzie, G. (1996). Regression models for survival data: The generalized time-dependent logistic family. *Statistician*, 45:21–34.
- MacKenzie, G. (1997). On a non-proportional hazards regression model for repeated medical random counts. *Statistics in Medicine*, 16:1831–1843.
- Morales, D., Pardo, L., Salicru, M. y Menendez, M.L. (1994a). Asymptotic properties of divergence statistics in a stratified random sampling and its applica-

Bibliografía

- tion to test statistical hypotheses. *Journal of Statistical Planning and Inference*, 38(2):201–224.
- Morales, D., Pardo, L. y Vajda, I. (1994b). Asymptotic divergence of estimates of discrete distributions. *Journal of Statistical Planning and Inference*, 48:347–369.
- Murphy, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Annals Statistics*, 22(2):712–731.
- Murphy, S.A. y van der Vaart, W. (1997). Semiparametric likelihood ratio inference. *The Annals of Statistics*, 25(4):1471–1509.
- Murphy, S.A. y van der Vaart, W. (2000). On profile likelihood. *Journal of the American Statistical Association, Theory and Methods*, 95(450):449–465.
- Pardo, L., Morales, D., Salicru, M. y Menéndez, M.L. (1993a). The ϕ -divergence statistics in bivariate multinomial applications including stratification. *Metrika*, 40:223–235.
- Pardo, L., Morales, D., Salicru, M. y Menéndez, M.L. (1993b). R_ϕ^h -divergence statistics in applied categorical data analysis with stratified sampling. *Utilitas Mathematica*, 44:145–164.
- Pardo, M.C. (1994a). On testing independence for multidimensional contingency tables with stratified random sampling. *Information Sciences*, 78:101–118.
- Pardo, M.C. (1994b). Test of independence for multidimensional contingency tables based on (h, ϕ) -divergence measures. *Utilitas Mathematica*, 48:75–91.
- Patilea, V. y Rolin, J.M. (2006). Product-limit estimators of the survival function for two modified forms of current-status data. *Bernoulli*, 12(5):801–819.

Bibliografia

- Pearson, K. (1990). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonable supposed to have arisen from random sampling. *Philosophy Magazine*, 50:157–172.
- Prentice, R. y Gloeckler, L. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34:57–67.
- Rao, J.N.K. y Scott, A.J. (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness-of-fit and independence in two-way tables. *Journal of the American Statistical Association*, 76:221–230.
- Read, T. y Cressie, N. (1988), *Goodness-of-fit statistics for discrete mutivariate data*. Springer Series in Statistics. Springer, New York.
- Royston, P. y Parmar, M.K.B. (2002). Flexible parametric proportional hazards and proportional odds models for censored survival analysis, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, 21:2175–2197.
- Samuelson, S.O. (1989). Asymptotic theory for non-parametric estimators from doubly censored data. *Scandinavian Journal Statistics*, 16:1–21.
- Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, 2:110–114.
- Solomon, H. (1960). Distribution of quadratic forms—tables and applications, Technical Report 45, Applied mathematics and statistics laboratories, Stanford University, Stanford, CA.
- Sun, J. (2005). Interval censoring. *Encyclopedia of Biostatistics*, Second Edition. John Wiley and Sons Ltd., 2603–2609.

Bibliografia

- Tsiatis, A.A. (1981). A large sample study of Cox's regression model. *The Annals of Statistics*, 9(1):93–108.
- Turnbull, B.W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical*, 69:169–173.
- van der Laan, M.J. y Gill, R.D. (1999). Efficiency of NPMLE in nonparametric missing data models. *Mathematical Methods of Statistics*, 8:251–276.
- van der Vaart, A.W. y Wellner, J.A. (1996). *Weak convergence and empirical processes: with applications to statistics*, Springer.
- van der Vaart, A.W. (1998). *Asymptotic statistics*, Cambridge University Press.
- Vaupel, J.W., Manton, K.G. y Stallard, E. (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16:439–454.
- Yang, S. (1991). Minimum Hellinger distance estimation of parameter in the random censorship model. *Annals of Statistics*, 19:579–602.
- Younes, N. y Lachin, J. (1997). Link-based models for survival data with interval and continuous time censoring. *Biometrics*, 53:1199–1211.
- Zografos, K., Ferentinos, K. y Papaioannou, I. (1990). φ -divergence statistics: sampling properties, multinomial goodness of fit and divergence test. *Communications in Statistics (Theory and Methods)*, 19(5):1785–1802.
- Zografos, K. (1993). Asymptotic properties of φ -divergence statistics and its application in contingency tables. *International Journal of Mathematics and Statistical Science*, 2(1):5–22.